



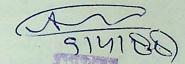
Digitized by Arya Samaj Foundation Chennai and eGangotri 14 3

Digitized by Arya Samaj Foundation Chennai and eGangotri CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar





110143



VOLUME LXXXV, NUMBER 1 JANUARY 1988

page

5 Biological Functions and Perceptual Content

Mohan Matthen

BOOK REVIEWS

- 28 Frederic Schick: Having Reasons: An Essay on Rationality and Sociality Edward J. Green
- 33 Gareth Evans: The Varieties of Reference Stephen Schiffer
- 42 David Lewis: On the Plurality of Worlds William G. Lycan
- 48 NOTES AND NEWS
- 53 MEMO TO AUTHORS
- 54 STATEMENT OF COPYRIGHT POLICY
- 55 JOHNSONIAN PRIZE IN PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Corliss Lamont; *Secretary*, Michael J. Mooney; *Treasurer*, Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

NOTE TO CONTRIBUTORS

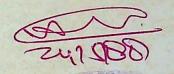
As a monthly periodical, THE JOURNAL OF PHILOSOPHY encourages philosophical exchange and progress through prompt publication of articles of current interest. The editors now feel that they can ensure publication within six to twelve months of acceptance. They continue to be particularly interested in papers that explore the borderline between philosophy and the special disciplines, but are unlikely to accept articles primarily expository or historical in character. For the Comments and Criticism section they accept rejoinders but prefer those designed to advance the discussion rather than just to set the record straight. If and only if articles are held five months or more and then rejected, letters of rejection include a brief comment concerning the editors' reasons for rejection.

Attention is drawn to the memo to authors and to the statement of copyright policy, republished on pages 53 and 54 of this issue; copies are available from the JOURNAL office.

Contributors are supplied free of charge with ten copies of the JOURNAL issues in which their articles appear, reviewers and authors of abstracts with two copies. Offprints may be ordered from the printer on a form supplied by the printer and enclosed with galley proof; they will be sent out within six weeks of publication.

© Copyright 1988 by the Journal of Philosophy, Inc.

ISSN 0022-362X



VOLUME LXXXV, NUMBER 2 FEBRUARY 1988

page

57 Logical and Analytic Truths that are not

Necessary

Edward N. Zalta

75 Second-order Logic Still Wild

Michael D. Resnik

COMMENTS AND CRITICISM

88 Burge's New Thought Experiment: Back to the Drawing Room

Kent Bach

98 NEW BOOKS

111 NEW BOOKS IN PAPERBACK

DIGITATE BY A SAME POUNDATION CHANNEL AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Corliss Lamont; *Secretary*, Michael J. Mooney; *Treasurer*, Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

Journal of Chinese Philosophy

Order form:
Dialogue Publishing Co.
P.O. Box 11071
Honolulu, Hawaii 96828 U.S.A.
Editor: Chung-ying Cheng
University of Hawaii, Honolulu USA
Co-Editor: Antonio S. Cua
Catholic University of America
Washington, D.C. USA
Editorial Assistant: Lauren Pfister

University of Hawaii, Honolulu, USA

INSTITUTIONAL SUBSCRIPTIONS
\$95 for one year
\$185 for two years
PERSONAL SUBSCRIPTIONS
\$39 for one year
\$73 for two years
Plus Postage:
AIR MAIL \$11.60 (per year/volume)
SEA MAIL \$6.60 (per year/volume)

Since its founding in 1973, the Journal of Chinese Philosophy has published over 150 series articles concerning subjects including classical Chinese philosophy, Neo-Taoism, Chinese Buddhism, Neo-Confucianism, and Chinese philosophy in the twentieth century, covering areas such as Chinese metaphysics, aesthetics, ethics and social-political philosophy. Symposium issues have been published on such themes as Chinese logic and Chinese philosophy of language, justice East and West, John Dewey and China, Chinese Cosmology and Metallurgy as well as Taoism and Anarchy.

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to Journal of Philosophy, 709 Philosophy, Columbia U, NYC 10027.

THE gitized by Arya Sahaj Floud Charles and the Gangotri OF PHILOSOPHY

2016101

VOLUME LXXXV, NUMBER 3 5119 1000 MARCH 1988

page		
115	Ducking Harm Christopher Boorse and	Roy A. Sorensen
135	Tachyons, Time Travel, and Divine Omni	science
		illiam Lane Craig
	BOOK REVIEWS	
151	Wayne Proudfoot: Religious Experience	Ninian Smart
154	Ross Harrison: Bentham	David Lyons
158	Michael Friedman: Foundations of Space-to	ime Theories:
	Relativistic Physics and Philosophy of S	
		Lawrence Sklar
164	Lewis Edwin Hahn and Paul Arthur Schilp	p, eds.: The
	Philosophy of W. V. O. Quine	Peter Hylton
160	NOTES AND NEWS	

FOUNDED BY FREDERICK I. E. WOODBRIDGE AND WENDELL T. BUSH Digitized by Arya Samaj Foundation Chennai and eGangotri

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00: students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. President, Arthur C. Danto; Vice President, Corliss Lamont; Secretary, Michael J. Mooney; Treasurer, Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell: and André Schiffrin.

o resec

AN INTERNATIONAL QUARTERLY OF THE SOCIAL SCIENCES

WINTER 1987

CONTEMPORARY CHINA

SOCIAL SCIENCE IN THE COMPARATIVE STUDY OF SOCIALISM IN CHINA: THE SOCIAL SCIENCES AT A CROSSROADS Gilbert Rozman

ANTHROPOLOGY IN THE PEOPLE'S REPUBLIC OF CHINA: THE WINDS OF CHANGE Gregory Eliyu Guldin

REFLECTIONS ON FAMILY RESEARCH IN THE PEOPLE'S NUREMBERG REPUBLIC OF CHINA

THE LEGACIES OF David Luban

Tamara K Hareven

MAO'S REPUBLIC David E Apter

HISTORIANS AND CON-SCIOUSNESS: THE MODERN POLITICS OF THE TAIPING HEAVENLY KINGDOM Robert P. Weller

Individual Subscriptions: \$20; Institutions: \$40 Single copies available on request Editorial and Business Office: Room GF354 66 West 12th Street, . New York, N.Y. 10011

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

ISSN 0022-362X

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to Journal of Philosophy, 709 Philosophy, Columbia U, NYC 10027.

CC-0. In Public Domain. Gurukul Kangri Collection, Handwal

THE Digitized by Anya and a Four Bring of Galgotri, OF PHILOSOPHY

VOLUME LXXXV, NUMBER 4

APRIL 1988

page

171 Human Nature and Social Change in the Marxist
Conception of History G. A. Cohen and Will Kymlicka

192 Moral Justification and Freedom Richmond Campbell

BOOK REVIEWS

214 Arthur C. Danto: The Philosophical Disenfranchisement of
Art and The State of the Art Alexander Nehamas

219 Jacques Derrida: The Truth in Painting David Carrier

223 ERRATUM NOTICE

224 NOTES AND NEWS

225 NEW BOOKS: TRANSLATIONS AND NEW EDITIONS

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Beroisky, oundation Chennyland, Senore Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. President, Arthur C. Danto; Vice President, Corliss Lamont; Secretary, Michael J. Mooney; Treasurer, Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

CANADIAN JOURNAL OF PHILOSOPHY SUPPLEMENTARY VOLUME 13, 1987

SCIENCE, MORALITY AND FEMINIST THEORY

Marsha Hanen and Kai Nielsen

Marsha Hanen Alison M. Jaggar

Annette C. Baier

Alison Wylie Sandra Harding Marilyn Friedman Virginia Held

Sibyl Schwarzenbach

John Exdell Kathryn Morgan

Steven Burns Barbara Houston

Susan Sherwin Christine Overall

Sheila Mullett Ann Ferguson

Lorraine Code Kai Nielsen

CDN \$14.00 (in Canada) US \$12.00 (outside Canada)
Please add \$1.50 for postage and handling,

and 50¢ for every further copy.

ISSN 0229-7051

Introduction: Toward Integration Sex Inequality and Bias in Sex Differences Research

The Need for More Than Justice The Philosophy of Ambivalence: Sandra Harding on The Science Question in Feminism Ascetic Intellectual Opportunities: Reply to Alison Wylie Beyond Caring: The De-Moralization of Gender

Non-Contractual Society Rawls and Ownership: The Forgotten Category of Reproductive Labour

Ethics, Ideology, and Feminine Virtue Women and Moral Madness

Moral Sanity or Who Killed Boy Staunton Rescuing Womanly Virtues: Some Dangers of Moral Reclamation Feminist Ethics and In Vitro Fertilization

Surrogate Motherhood Only Connect: The Place of Self-Knowledge in Ethics A Feminist Aspect Therapy of the Self Second Persons

Afterword: Feminist Theory - Some Twistings and Turnings

This Supplementary Volume is free to individual and Student SUBSCRIBERS to CJP Volume 17, 1987.

ORDER FROM:
THE UNIVERSITY OF CALGARY PRESS, LT 1013
THE UNIVERSITY OF CALGARY
2500 UNIVERSITY DRIVE N.W.,
CALGARY, ALBERTA T2N IN4 CANADA

ISBN 0-919491-13-8

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to Journal of Philosophikul RangPhilosophion, Chardenna U, NYC 10027.

VOLUME LXXXV, NUMBER 5 MAY 1988

page 227 253		Jerrold J. Katz en R. Leighton
265 271	BOOK REVIEWS Alvin Goldman: Epistemology and Cognition Colin McGinn: Wittgenstein on Meaning	Fred Dretske T. W. Child
277 279	NOTES AND NEWS NEW BOOKS: ANTHOLOGIES	

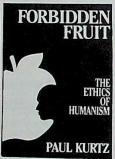
Digital Eby Mod & Roll Adundant Creatil Line Son Port Y

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Corliss Lamont; *Secretary*, Michael J. Mooney; *Treasurer*, Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.



FORBIDDEN Forbidden Fruit The Ethics of Humanism

Paul Kurtz

Is there a school of ethics that may be genuinely called "Humanistic?" Paul Kurtz says "yes"—and, in this book, he attempts to work out both its theoretical foundations and its normative implications.

There are chapters on the "common moral decencies," "the ethics of excellence," responsibility, human rights, moral education for children, privacy, the meaning of life, and the ethics of the world community.

275 pages

ISBN 0-87975-454-0 • Cloth \$19.95

ISBN 0-87975-455-9 • Paper \$11.95

Prometheus Books

700 East Amherst Street, Buffalo, New York 14215

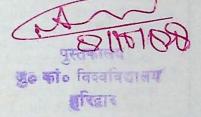
The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to Journal of Philosophy, 709 Philosophy, Columbia U, NYC 10027.

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE JOURNAL OF PHILOSOPHY



VOLUME LXXXV, NUMBER 6
JUNE 1988

page		
283	Ethical Universalism and Particularism	Alan Gewirth
303	Apportioning Causal Responsibility	Elliott Sober
319	Personal Identity and Survival	John Robinson
	BOOK REVIEWS	
329	Graeme Forbes: The Metaphysics of Modal	lity
		Stephen Yablo
338	NOTES AND NEWS	

THE JOURNAL COUNTRICOS COPPY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Corliss Lamont; *Secretary*, Michael J. Mooney; *Treasurer*, Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY FOUNDED BY MARVIN FARBER ERNEST SOSA, Editor RODERICK M. CHISHOLM, Associate Editor

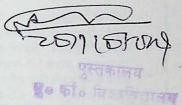
PPR publishes articles in a wide range of areas including philosophy of mind, epistemology, ethics, metaphysics, and philosophical history of philosophy. No specific methodology or philosophical orientation is required in submissions. An abstract not exceeding 150 words and two nonreturnable copies of any manuscript submitted for publication are required.

Address All Communications to: Philosophy and Phenomenological Research BROWN UNIVERSITY, BOX 1947 PROVIDENCE, RHODE ISLAND 02912

Annual subscription rates: \$20.00 (foreign, \$23.00) for libraries and institutions \$15.00 (foreign, \$17.00) for inditional charge for postage and handling on all back issues. A circular listing the main contents of the journal and a list of available back issues will be sent upon request.

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal liof Danielop Ryrukul Kangri Collection, Haridwar ISSN 0022-362X POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices.



VOLUME LXXXV, NUMBER 7 JULY 1988

page			
339	The Return of the Gene	Kim Sterelny and Phi	lip Kitcher
362	Coping with Conflict	Fred	eric Schick
		\$	
	BOOK REVIEWS		
376	Martha Nussbaum: The Fr	ragility of Goodness T	. H. Irwin
384	Jerry Fodor: Psychoseman	tics Daniel	C. Dennett
390	NEW BOOKS		

Digitized by Apa Sairan Auhdato Cherill III as OPiHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Corliss Lamont; *Secretary*, Michael J. Mooney; *Treasurer*, Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

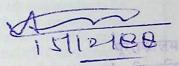
PHILOSOPHICAL INQUIRY

An International Philosophical Quarterly
Adress: Aristotelian University of Thessaloniki
POBOX 84, Thessaloniki, Greece
The last copies of the following special
issues are now available:
Aristotle \$ 10, Plato \$ 10, Aesthetic
Criteria \$ 10, Ecophilosophy \$ 10

D. Z. Andriopoulos:
THE CONCEPT OF CAUSALITY
IN PRESOCRATIC PHILOSOPHY,
«Philosophical Inquiry» Monographs,
(1988), Pgs 132, price \$ 15.00
Orders: Publishing House VANIAS, 26 Armenopoulou, Street, Thessaloniki, Greece

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.
POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices.
Send address changes to Journal of Philosophy, 709 Philosophy, Columbia U, NYC 10027.



VOLUME LXXXV, NUMBER 8 AUGUST 1988

395 Sanctification, Hardening of the Heart, and Frankfurt's

Concept of Free Will Eleonore Stump

COMMENTS AND CRITICISM

- 421 Free Riding and the Prisoner's Dilemma Raimo Tuomela
- 427 Metaphor without Mainsprings: A Rejoinder to Elgin and Scheffler Josef Stern
- 439 NOTES AND NEWS
- 440 NEW BOOKS
- 444 NEW BOOKS: ANTHOLOGIES
- 447 NEW BOOKS: TRANSLATIONS AND NEW EDITIONS

Published by The Journal of Philosophy, Inc.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

THE JOURNAL OF PHILOSOPHY Digitized by Arya Samaj Foundation Chennai and eGangotri FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. President, Arthur C. Danto; Vice President, Corliss Lamont; Secretary, Michael J. Mooney; Treasurer, Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

PHILOSOPHY EAST & WEST A Quarterly of Asian and Comparative Thought

Forthcoming articles

Exploring the Confucian Tradition Wing-tsit Chan

A Fragment of the Indian Philosophical Tradition—Theory of Pramana J. N. Mohanty

The Buddhist Conceptions of "Subject" and "Object" and their Moral Implications David J. Kalupahana

Shang Divination and Metaphysics David N. Keightley

The Problem of Conceptual Unity in Hsün Tzu and Li Kou's Solution Antonio S. Cua

The Philosophical Concept of Foreknowledge in the Thought of Shao Yung

Anne D. Birdwhistell

Institutions, 1988 rates—single copy, \$7; one year, \$21 Individuals, 1988 rates—single copy, \$5; one year, \$16 Students, 1988 rates—single copy, \$5; one year, \$10

UNIVERSITY OF HAWAII PRESS

2840 Kolowalu Street, Honolulu, Hawaii 96822 U.S.A.

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

POSTMASTERCS Cond Rubbic Postage Pallukul Kangti Collection, Haridwar Send address changes to Journal of Philosophy, 709 Philosophy, Columbia U, NYC 10027.

P

LISRASY

VOLUME LXXXV, NUMBER 9
SEPTEMBER 1988

page

451 Justice, Care, Gender Bias

464 An Explanation of Retribution

479 Motive and Justification

BOOK REVIEWS

500 Jerrold J. Katz: Cogitations

506 NOTES AND NEWS

Cheshire Calhoun

Andrew Oldenquist Stephen Scott

Palle Yourgrau

700

DIGREFFEDY POISPON ADLINGTOFF CIPHINI IN OSCIPHINY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. President, Arthur C. Danto; Vice President, Corliss Lamont; Secretary, Michael J. Mooney; Treasurer, Barbara Gimbel; Leigh S. Cauman; Stephen S. Fenichell; and André Schiffrin.

REVUE INTERNATIONALE DE PHILOSOPHIE

Editor: Michel MEYER 143, av. A. Buyl, 1050 Brussels, Belgium

Each number is devoted to a particular movement, a particular philosopher, or a particular problem.

We publish 3 issues annually. Articles are written in English, French, German, or Italian.

Coming numbers

Locke
Kant and Practical Reason
Artificial Intelligence
Heidegger
Wittgenstein

Our last issue was devoted to *Schopenhauer* (articles by D. W. Hamlyn, D. E. Cartwright, W. Schirmacher and K. Boullart)

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

POSTMASTER: Second Public Domain. Gurukul Kangri Collection, Haridwar ISSN 0022-362X

Send address changes to Journal of Philosophy, 709 Philosophy, Columbia U, NYC 10027.

THE Digitized by Arma Farration Chenna and edungotri O.F. OHICOSOPHY

VOLUME LXXXV, NUMBER 10

OCTOBER 1988 Gurukul Kongri Vistavara 6

Special Issue, Eighty-fifth Annual MeetingHARBWAR

American Philosophical Association, Eastern Division

F	American Philosophical Association, Laste	The Division
page	Symposium: The Philosophical Significance Experimentation	of 2311
507	On the Stability of the Laboratory Sciences	Ian Hacking
515	Experiment and Theory: Constitution and	Reality
	I	Patrick A. Heelan
525	Philosophy in the Laboratory	Peter Galison
	Symposium: Postmodernism and Rationality	
528	Postmodernism and Rationality	Rodolphe Gasché
539	The Question of Reason	Charles Scott
	Symposium: Plato and the Proleptic Readin Dialogues	ng of Socratic
541	Plato's Charmides and the Proleptic Readi Dialogues	ing of Socratic Charles H. Kahn
550		s L. Griswold, Jr.
552	Symposium: Argument and Perception Argument and Perception: The Role of Lin Moral Inquiry	terature in Iichael R. DePaul
566	Moral Theory and Moral Education	R. B. Brandt
569	NOTES AND NEWS	
571 583 588 600	Program of the Meetings List of Program Participants Appendix: Programs of Group Meetings List of Group Participants	
- 0 0	List of Group ratherparts	

Dig TOP AND SARANT ALLICANO DE NERTHAL ILLO CASO PRIMY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro. Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. President, Arthur C. Danto; Vice President, Corliss Lamont; Secretary, Michael J. Mooney; Treasurer, Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

Noûs

Prof. Hector-Neri Castañeda, editor, Indiana University Prof. Earl Conee, associate editor, University of Rochester

The primary purpose of Noûs is to publish outstanding essays that extend the frontiers of philosophical research. Noûs emphasizes positive theoretical work that takes full account of current developments in philosophy. There are no editorial restrictions on either the area of philosophy that is discussed or the philosophical orientation of the discussion. Noûs offers several distinctive services to its readers:

CRITICAL REVIEWS: Each issue contains detailed evaluations of a few important recent philosophical books.

CENTRAL DIVISION SYMPOSIA PAPERS: The March issue of each volume is devoted to the invited addresses for the year's American Philosophical Association Central Division Meetings.

SPECIAL TOPIC ISSUES: Frequently an issue will contain a collection of major papers on a single central philosophical topic, often guest-edited by a leading authority on the topic.

Subscription Rates for Volume XXII (1988) only: Individuals: 1 year \$22.00, 2 years \$40.00, 3 years \$60.00; Institutions: 1 year \$45.00, 2 years \$90.00, 3 years \$135.00. Please add \$5.00 postage per year for all foreign subscriptions.

INQUIRIES SHOULD BE SENT TO: Noûs, Department of Philosophy, Indiana University Sycamore 126, Bloomington, Indiana 47405 USA.

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

Pul

POSTMASTER: Second Platsid Pulagenp Edruk W Kanggirk, of Pertian Haridwar Send address changes to Journal of Philosophy, 709 Philosophy, Columbia U, NYC 10027.

THE JOURNAL CHEMAIN CH

VOLUME LXXXV, NUMBER 11 NOVEMBER 1988

Special Issue, Eighty-fifth Annual Meeting American Philosophical Association, Eastern Division

page	Symposium: What Is Still Valuable in Huss	erl's Analyses of
	Inner Time-consciousness	
605	What Is Still Valuable in Husserl's Anal	yses of Inner
003	Time-consciousness P	eter K. McInerney
617	Comment on McInerney	J. Claude Evans
	Symposium: Feminist Skepticism	
619	Feminist Skepticism and the "Maleness"	of Philosophy
0.10	Tomorous Stopheses and	Susan Bordo
630	Further Thoughts on a "Theoretics of He	terogeneity"
000	Tarther Thoughts on a Theoretic sy	Naomi Scheman
	Symposium: Law and Society	
632	The Politics of Friendship	Jacques Derrida
645	On the Margins of Politics Th	omas A. McCarthy
	Symposium: Individuation and Self-knowle	edge
649	Individualism and Self-knowledge	Tyler Burge
664	Reply to Burge	Donald Davidson
	Symposium: History of Philosophy as a Di	scipline
666	The History of Philosophy as a Disciplin	
673	The History of Philosophy as Actual Phi	
	22000 y of 1 helosophy as 2200aut 2 he	Kenneth Schmitz

Digitiz**FqHyP**ary**F@HJRFNAda**tio@FendalHhLloOsG@PHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. *President*, Arthur C. Danto; *Vice President*, Corliss Lamont; *Secretary*, Michael J. Mooney; *Treasurer*, Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.



VOLUME XX:1-2 Fall-Winter 1988-89

\$7.50

LATIN AMERICAN A SPECIAL DOUBLE ISSUE

Horacio Cerutti-Guldberg Jorge J.E. Gracia Iván Jaksić Adolfo Sánchez Vázquez Ofelia Schutte David Sobrevilla Leopoldo Zea Editor: Marx Wartofsky

Guest Co-Editor: Jorge J.E. Gracia

SUBSCRIPTIONS \$15 per year for individuals \$45 per year for institutions + \$4 for foreign subscriptions

ICAN PHILOSOPHY TODAY

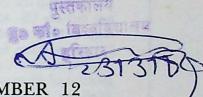
Also VOLUME XIX:2-3 Winter-Spring 1988 FICHTE and Contemporary Philosophy \$7.50

the philosophical forum PO. Box 239 Baruch College of CUNY 17 Lexington Avenue New York, NY 10010

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

POSTMASTER Good In Public Domain. Gurukul Kangri Collection, Haridwar ISSN 0022-362X Send address changes to Journal of Philosophy, 709 Philosophy, Columbia U, NYC 10027.



VOLUME LXXXV, NUMBER 12 DECEMBER 1988

675 Virtue Concepts and Ethical Realism Bruce W. Brower
694 Rules for a Good Language Eli Hirsch

718 A Note on Categorical Properties and Contingent Identity

Jerrold Levinson

723 NOTES AND NEWS

725 INDEX TO VOLUME LXXXV

Dightzed by Jaya Samaj Foundation Chennai and eGangotri Founded by Frederick J. E. WOODBRIDGE AND WENDELL T. BUSH

Editors: Bernard Berofsky, Arthur C. Danto, Hidé Ishiguro, Isaac Levi, Sidney Morgenbesser, Mary Mothersill, Charles D. Parsons, Gisela Striker, James J. Walsh. Consulting Editors: Akeel Bilgrami, James T. Higginbotham, Charles Larmore. Managing Editor: Michael Kelly. Editorial & Business offices: 709 Philosophy Hall, Columbia University, N.Y.C. 10027.

All communications to the editors, manuscripts, advertisements, subscriptions, and changes of address should be sent to the Managing Editor. Annual subscription price (12 numbers) in US currency on a US bank: individuals, \$25.00; institutions, \$40.00; students and the retired and unemployed, \$20.00; plus \$6 postage outside US. All back volumes and separate issues available back to 1904. Published monthly as of January 1977; printed by Lancaster Press, Inc., Lancaster, Pennsylvania.

THE JOURNAL OF PHILOSOPHY is owned and published by The Journal of Philosophy, Inc. President, Arthur C. Danto; Vice President, Corliss Lamont; Secretary, Michael J. Mooney; Treasurer, Barbara Gimbel; Leigh S. Cauman; and André Schiffrin.

STATEMENT OF OWNERSHIP, M	Tel Service ANAGEMENT AND CIRCULA 9 U.S.C. 3645)	TION
JOURNAL OF PHILOSOPHY	18. PUBLICATION NO.	2. Date of Filing 2 X 8/2/88
9. For Completion by Nonprofit Organizations Authorized To Mail at Special The purpose, function, and nonprofit status of this organization and the research of the purpose, function, and nonprofit status of this organization and the research of the purpose of	exempt status for Federal income tax purp	iblisher must submit explanation of
10 Extent and Nature of Circulation (See instructions on reverse side)	Average No. Copies Each Issue During Preceding 12 Months	Actual No. Copies of Single Issue Published Nearest to Filing Date
A Total No Copies (Net Press Rum)	4845	4737
Paid and/or Requested Circulation Sales through dealers and carriers, street vendors and counter sales		
2. Mad Subscription (Paid and/or requested)	3961	3908
C Total Paid and/or Requested Circulation (Sum or 1081 and 1082)	3961	
D Free Distribution by Mail, Carrier or Other Means Samples, Complimentary, and Other Free Copies	174	3908
E Total Distribution (Sum of C and D)		170
F Copies Not Distributed 1 Office use, left over, unaccounted, spoiled after printing	4135	4086
2 Return from News Agents	710	651
G. TOTAL ISum of E. Fl. and 2-should equal net press run shown in A)		
	4845 Entr Of Edgio Publishin, Burnasa Mana	4737 ger, or Owner
Form 3526, Dec. 1997	I Kelly, Managing	Editor

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

POSTMASTER Committed Second Relation Postagen particular Kangri Collection, Haridwar Send address changes to Journal of Philosophy Took, NY, and additional mailing offices. Send address changes to Journal of Philosophy, 709 Philosophy, Columbia U, NYC 10027.

914100

THE JOURNAL OF PHILOSOPHY

THE JOURNAL OF PHILOSOPHY was founded in 1904 as THE JOURNAL OF PHILOSOPHY, PSYCHOLOGY AND SCIENTIFIC METHODS by Frederick J. E. Woodbridge and J. McKeen Cattell. In 1906 Wendell T. Bush became associated with the JOURNAL as co-editor. In 1923, the JOURNAL was incorporated in the State of New York under its present name. Since its founding it has been published from Columbia University and printed by the Lancaster Press in Lancaster, Pennsylvania.

VOLUME LXXXV JANUARY-DECEMBER 1988

Edited by

Bernard Berofsky, Arthur C. Danto Hide Ishiguro, Isaac Levi, Sidney Morgenbesser Mary Mothersill, Charles D. Parsons Gisela Striker, James J. Walsh

The appearance of the code at the bottom of the first page of an article in this JOURNAL indicates the copyright owner's consent that copies of the article be made for personal or classroom use. This consent is given on condition that the copier pay the stated per-copy fee through the Copyright Clearance Center, Inc., P.O. Box 765, Schenectady, New York 12301, for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying.

© Copyright 1988 by the Journal of Philosophy, Inc.

POSTMASTER: Second class postage paid at New York, NY, and additional mailing offices. Send address changes to Journal of Philosophy, 709 Philosophy, Gallundaia U, NYC 10027.

VOLUME LXXXV, NO. 1, JANUARY 1988

BIOLOGICAL FUNCTIONS AND PERCEPTUAL CONTENT*

HE computational approach to perception has become increasingly prominent over the last decade or so. The recent publication of David Marr's *Vision*¹ makes computational theories accessible to philosophers, who may now judge for themselves how unified, wide–ranging, powerful, and intellectually satisfying the program has proved to be.

A striking feature of computational theories of perception, from the philosophical point of view, is that they employ the intentional characterizations of perceptual states insisted upon by Edmund Husserl and other phenomenologists. Most other scientific treatments of perception concentrate on the external causes of perception and how they result in the stimulation of the sensory organs, the neurological organization of perceptual systems, and the discriminatory resources afforded to organisms by the perceptual systems. Though not ignoring these important topics, the computationalists characteristically regard perceptual states as representing the external world and seek to explain them under this rubric. Marr claims, for example, that vision represents three-dimensional surfaces as composed of "generalized cones," and tries to show how such representations may be derived from representations of two-dimensional outlines of the sort found in line drawings and how the latter may, in turn, be extracted from a pattern of optical stimulation present on

Most philosophers today, and most computational theorists, are materialistically inclined: they are not inclined to allow that there

¹ New York: Freeman, 1982.

^{*} A version of this paper was read to the philosophers at Dalhousie University. I am grateful to them all for lively discussion, and to Terry Tomkow and Bob Martin in particular for helpful suggestions. I owe no small debt to the perception reading group in my department, and especially to Bruce Hunter.

Digitized by Arya Samaj Foundation Chennai and eGangotri

could be a causally effective aspect of a perceptual state that is not material.2 For this reason, any widely acceptable foundational account of computational theories of perception must explicate what it means to endow a material state with representational content. It is my purpose in this article to develop a theory of perceptual content which meets this requirement.

An important constraint on such a theory results from the fact that computational theories tend to attribute distal content to perceptual states. They claim that we perceive things like rigid bodies, motions in three-dimensional space, and reflectances of surfaces: these are entities in the external world. Thus, the theory we are looking for cannot identify content with some purely internal concomitant of a perceptual state—it is important, in other words, to avoid the materialistic analogue of sense-datum theories, which claim that the content of a state is properly specified by the introspectible feel that accompanies it.

It is often assumed that the alternative to "internalist" theories of perceptual content is a causal theory. I shall argue in section III that this too is a false lead. Sections II and III prepare us for a new approach to perceptual content. The theory of perceptual content I shall propose is based on biological functions. It employs a notion of biological adaptation which results from an extension of Larry Wright's work on functions (sections IV and V). I shall show (section VI) how chains of adaptations are useful in accounting for the attribution of content, first, to states that may be regarded as the evolutionary precursors of perception and, then, to the full-blown perceptual states that humans experience (section VII). I shall argue that the content-attribution scheme I propose accounts for distality of content and, indeed, that it can form the basis of realism with regard to perception (section VIII). Finally, I shall make a suggestion how the content-attribution scheme might be able to account for the point and utility of intentional attribution in general (section IX).

I. WHAT IS PERCEPTUAL CONTENT?

According to Husserl, a perceptual state is about something, its object of perception. It achieves this aboutness by means of its noema. According to one influential interpretation, a noema stands to the object of perception much as the Fregean sense of a referring expression stands to its reference.3 This conceptualization of the con-

² Qualia seem to constitute an exception: many psychologists talk as if they exist and influence our actions.

³ The interpretation is due to Dagfinn Føllesdal: see his "Husserl's Notion of Noema," this JOURNAL, LXVI, 20 (Oct. 16, 1969): 680-689. The classic source of Husserl's own exposition is his *Ideas* (1913), W. R. Boyce Gibson, trans. (London:

tent of a perceptual state is, I believe, particularly appropriate to explaining the structure of computational theories. Let us say that the content of a perceptual state, its Husserlian noema, is expressible by means of a definite description. Then the object of perception is the unique object that satisfies the definite description. If there is no such object, the perception is nonveridical—it is a misperception.4

The content of a perceptual state may be regarded as having two components. First, it presents an object as possessing a certain characteristic-it may present an object as red, for example, or as moving in a certain direction, or as five feet away. Second, there is something about the state which accounts for the fact that the state has a unique object—not just any red thing is the object of a redpresenting state, but some unique red thing. An adequate theory of perceptual aboutness must account for both these features of perception, its mode of presentation and its individual-directedness.

In my opinion, any adequate theory of the individual-directedness of perception will be causal. A promising theory along these lines has recently been offered by David Woodruff Smith, elaborating on an idea of John Searle's.5 According to this theory, the definite description that expresses the content of a perceptual state has a self-referential causal-indexical component. Searle's idea is that the content of a perceptual state is always expressible in this form: the individual that has the property F and is causing THIS state. The object of a perceptual state, then, will be whatever satisfies this description.

The only variable element in content so expressed is the property F by which the individual is characterized. Every perceptual state is directed toward some individual causing the state itself, but some states are directed toward the red object causing it, some toward the moving object causing it, and so on. This variable element is the mode of presentation. Computational theories are efforts to explain how perceptual states present external objects. The question I shall seek to address in what follows is this: What feature of a perceptual state makes it a presentation of a particular property, say redness? As a warning against a possible confusion, I should remark that, even though content as a whole has a causal-indexical element, it does not

⁵ Searle, Intentionality (New York: Cambridge, 1983), pp. 61-71; and Smith, "The Ins and Outs of Perception," Philosophical Studies, XLIX, 2 (March 1980):

not

ac-

at it

lt is

ent

hat

ual

ons

are

for

of a

ite-

on-

hat

of

nat

it I

of

rry

on

ri-

lu-

er-

ıat

of

rd

he

nt

b-

a.ne X-

n-

ist

of

of

n:

⁴ This neglects what David Lewis has called "veridical hallucinations": see his "Veridical Hallucination and Prosthetic Vision," Australian Journal of Philosophy, LVIII (1980): 239-249. This is a problem that concerns the formulation of what I shall call the individual-directedness component of the noema, not the mode of presentation, and so it will not concern us in this paper.

Digitized by Arya Samaj Foundation Chennai and eGangotri follow that an account of the mode of presentation should be causal as well.

II. INTERNALIST CRITERIA OF CONTENT

Some epistemologists have claimed that the quale, or subjective character of a state, constitutes its mode of presentation—thus a state presents an object as red in virtue of its being associated with a red quale or sense datum. Others give some internal material state the same role as qualia. Just as pain has been identified with a stimulation of the C-fibers, a red-presentation might be identified with some neurological concomitant of seeing red. Such theories are *internalist* because they ascribe content to states solely on the strength of their intrinsic character and refer to nothing outside the perceiver, especially not to the external things represented by perception.

An internalist theory is unsatisfactory for several reasons. Most importantly for our present purposes, it cannot adequately account for the distality of perceptual content found in computational theories. As an example of what I mean, let us consider in some detail Edwin H. Land's theory of color vision, which is regarded by many as an early paradigm of the computational approach to perception.

The cone cells of the retina are sensitive to the color of light. Consequently, many psychologists and many philosophers have assumed (in a tradition going back to Newton's *Optics*) that the color we see is the color of the light incident on our retinas. But this is false. Our perceptions of color are, in fact, much more closely correlated with the true color of surfaces, even when they reflect light that might be expected to be quite misleading. This was recognized to be the case as long ago as Helmholtz, but most conclusively established in a series of experiments, conducted by Land and J. J. McCann, in which human observers were shown to be able to detect that two surfaces were different in color, and what their true colors were, even though the illumination was so rigged that the surfaces were sending light of exactly the same color to the eye and so should have been indistinguishable.

But how can vision detect true color if it is sensitive only to light color? The color of light reflected from a surface is a function of two independent variables—the color of the surface itself, and the color of the illumination incident upon it.⁷ Thus, the color of light incident

⁶ "The Retinex Theory of Color Vision," Scientific American, CXXXVII, 6 (December 1977): 108–128; and Land and J. J. McCann, "Lightness and Retinex Theory," Journal of the Optical Society of America, LXI (1971): 1–11. Marr (17) tion.

⁷ Let us divide visible light into three wavelength ranges: long, medium, and short. Then the color of a surface can be defined as the ordered triple of the

upon the retina is compatible with an infinite number of covarying combinations of surface color and illumination color. What is needed, then, from an adequate theory of color vision, is an account of how the visual system is able to discriminate surface colors, given only light color; in other words, how it separates out the contributions to afferent light color due to each external factor.

sal

ar.

ate

ed

he

on

ne

al-

of

er,

ost

nt

al

ail

ny

n.

ıt.

ıs-

or

is

r-

ht

ed

b-

ct

rs

es

d

d

In the world in which we live, there are many abrupt changes in surface color, but very few abrupt changes of illumination. According to Land, color vision uses a process that is effective because of this fact. It distinguishes types of brightness gradient in the image incident upon the retina. Sudden changes of brightness are identified as being due to changes in surface color, and more gradual changes are identified as being due to changes in illumination. In this way, our vision assigns a surface—color *gradient* to each point in a scene. *Absolute* colors are then assigned to points by a process of calibration, the details of which need not concern us here.

It is very important to be clear on what Land uses to explain what. First, Land uses psychological experiments to identify a physical variable, surface color, as that which is presented by color vision. The problem he then tackles is that information concerning this detected property is unavailable at the retina. So Land appeals to an *indicator* of the detected property which he claims *is* present in the retinal image, and uses it to explain how the visual system has access to the presence of the external property.

Obviously this procedure raises a number of interesting questions. What is important for our immediate purposes is that Land type—identifies perceptual states by the external property they present, not by their quale nor by any other internal characteristic. The direction of explanation in Land's theory is from external properties and their indicators to the content of states. Internalist theories reverse this direction of explanation. For them, the similarities relevant to content attribution are internal, not external, and any representational relationship borne by states to external reality must be accounted for by internally specified manipulations of content. As an extreme of this attitude, consider qualia theories. They type—identify states by their qualia, and identify external objects as having the same property because they tend to produce qualitatively similar states in us.⁸

reflectances of the surface for light in these three wavebands. The color of a beam of light is the average of wavelengths in the beam weighted by the energy present in a surface of wavelength.

⁸ There are two versions of a definition of color by means of qualia. The first sort makes color a construction out of qualia. The second sort uses qualia to identify a

Now, it is possible to argue that Land's type-identification will have to be nearly coextensive with a type-identification by internal criteria. For, as we have seen, the experimental verification of Land's hypothesized mode of presentation lay in the ability of observers to identify the true colors of surfaces (even when they sent misleading signals to the eye). One would think that observers identify a surface as having a color as a result of some internally available feature of the perception, perhaps a quale. So there must be some intrinsic difference between perceptions of different things. To put this point in general terms, a perceiver's epistemic access to the content of his or her perceptual states must be through internal characteristics. This access has to be reliable, if not infallible; so there will always be an extensionally correct (or nearly correct) internalist definition of content.9 Nevertheless, such definitions will misrepresent computational theories, because they will miss what such theories take to be the essence of content.

III. CAUSAL THEORIES OF THE MODE OF PRESENTATION

There are two sorts of causal theory. The first kind, which I shall call outside-in, attributes content to a perceptual state on the basis of what causes it. Broadly speaking, a theory of this type will allow that two states with the same intrinsic (i.e., internally specifiable) character may have different content, because two states with the same character may have different causal origins.

The shortcomings of the outside—in approach have been decisively exposed by Jerry Fodor. ¹⁰ If a visual state counts as being of a rigid body simply because it was caused by one, then the content of the state will be unhelpful in explaining the use a perceiver makes of it. A visual state might have been caused in me by a rigid body without that fact being known to me. My treating something as a rigid body must, then, be explained by invoking an additional state, my perceiving it as a rigid body. But, if the latter state is invoked, the fact that the

real property. [An example of the second sort: red is the external property that causes this sort of quale in me. For an ingenious discussion of this sort of definition, see Edward Wilson Averill, "Color and the Anthropocentric Problem," this JOUR-NAL, LXXXII 6 (June 1985): 281–304.] It is the first sort of definition which is criticized in this section. The second sort is criticized, in effect, in the discussion of Keith Campbell and C. L. Hardin in section VIII below.

⁹ Some philosophers would contest the value of such a definition, however. It does not follow from my argument that there will be *inter-personally applicable* purports to show such a definition false is the inverted-spectrum argument. There is an obvious materialistic analogue of this argument: presumably, token-token identications are such as the inverted of the content of t

^{10 &}quot;Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology," Behavioural and Brain Sciences, III, 1 (1980): 63-73.

former is or is not caused by a rigid body will be otiose—what I do will be the same in either case.

will

nal

id's

to

ing

ace

the

er-

in

or

his

an

of

ta-

be

all

of

nat

ac-

ne

ely

rid

he

A

at

st,

it

ne

at

n,

Ris

of

It le

at

is

1-

e

An *inside-out* theory of perceptual content recognizes that there must be internal features of a state which provide the perceiver with access to its content. It thus looks at intrinsic features of a perceptual state and asks: How does a state with these intrinsic features present its object? In this way, it assigns distal content to *internally identifiable types* of states.

Now, suppose that we have two states with the same intrinsic character, one caused by a particular external object or feature, and the other not. An inside—out theory will assign the same content to both states, and so this assignment must not imply that the external object or feature is present. There are two options available for achieving an assignment of this sort.

First, one may retain the assumption that content is cause, and so retreat to some more proximal common cause of the two states. For example, when it was believed that two different surface colors could result in the same sort of state, under different conditions of illumination, philosophers claimed that it is not surface color that we see, but rather the color of the light incident upon the eye. Similarly, it has been held that what we see is projected shape, size, and motion, not actual shape, size, and motion—the move is familiar. It is this move which we wish to avoid in order to maintain the characteristic distality of content attribution in computational theories.

The other alternative is to drop the causal assumption. This approach need not be as radically at odds with the general causal intuition as it looks, for one might say that a state with content *C* is usually or normally caused by an object that is *C*. My argument here seeks to show that no such move will work. Causal theories of the mode of presentation have to be abandoned altogether.

Let us look at the matter in more detail. We have seen that an inside—out theory of perceptual content attempts to assign distal content to internally identifiable types of states. If a causal theory is seen as the way to account for distality, an inside—out theorist will have to define content in terms of what causes a state of a certain sort. There are two versions of such a theory, depending on whether we define content by the *invariable* or the *normal* causes of a kind of state:

IOC: A perceptual state with (internal) feature S will have mode of presentation F if S-states are invariably/normally caused by things that have the property F.

It is a familiar fact that theories of content which appeal to the invariable causes of kinds of states cannot deal with misperception.

For, if a state that presents a thing as F is always caused by a thing that is F, then it cannot come into being unless it is veridical. I shall argue that theories that appeal to the normal causes of misperception will not work either: for they cannot deal with what I shall call normal misperception.

Before we find out what normal misperception is, let us briefly look at two sorts of misperception which are adequately accounted for by the version of IOC which appeals to the normal causes of a sort of perceptual state. The first is caused by malfunction: if a system is not working normally, because of genetic damage or injury, then misperceptions may result. Second, maladaptation: if a system is exposed to environmental circumstances to which it is not adapted, circumstances manufactured by a laboratory worker for example, then it may be "fooled." Is misperception always due to malfunction or maladaptation, as Fodor conjectured in "Semantics, Wisconsin Style"? If so, then, on a plausible understanding of the 'normally' in IOC, we should have an adequate causal account of the mode of presentation.

But there is another source of misperception. Perceptual states play a mediating role between external circumstances and appropriate behavior. But the features of the external world which affect a system may not be those which are "task–relevant." This is immediately apparent in the case, already described, of color vision. Our retinas are sensitive to the color of light. But light color tells us little that is useful about the things we have to deal with; so we are in need of a system that is able to process afferent "sensory information," yielding a characterization of the external world in terms of surface color.

Now, in carrying out this "processing," the visual system is hampered by the fact that the afferent sensory information it works with underdetermines conclusions about surface color. That is, the afferent information is compatible (given the laws of how light is reflected by surfaces) with several different and incompatible "conclusions" about surface color. So color vision uses an indicator of surface color which is present in the retinal image to arrive at the representation the organism needs. However—and this is the important point—the indicator is not perfect. It is possible to create a surface on which the color changes gradually, in just the way that illumination normally changes. In uniform illumination, such a surface would be perceived as uniformly colored and variably illuminated. Here we have a

¹¹ See Fodor, "Semantics, Wisconsin Style," Synthese, LIX, 3 (June 1984):

¹² Imagine a white table-cloth illuminated from one end by a white lamp and from the other by a red lamp. The illumination will want our additional control of the control

13

situation in which the visual system represents a surface-color gradient, nonveridically, as an illumination gradient; in other words, we have a misperception.

ing

hall

ep-

call

efly

ted

of a

fa

ıry,

em

ed.

ole,

ion

sin

in.

of

tes

ro-

t a

di-

ur

tle

ed

1,"

ır-

m-

th

er-

ed s"

or

on

ne

ne

ly

ed

):

nd

The kind of misperception I want to draw attention to arises from the fact that our ability perceptually to discriminate the colors of surfaces in widely varying conditions of illumination is made possible by the use of indicators that are imperfect, but nevertheless the best available. If the visual system did not use such indicators, the only kind of color we perceived would be the color of light incident upon the retina, which is useless. The price of using the imperfect correlation is occasional misperception, a small price to pay for a useful representation of the world.

It is obvious that the misperceptions arising from the use of such imperfect indicators will not be traceable to *malfunction*: the error lies in the indicator itself, not in how it is used. Nor can the misperception be claimed to result from the visual system's being *maladapted* to the situations it encounters; for the indicator is not perfectly reliable even in the range of situations to which the visual system is adapted. (In many such cases, no indicator could be, since the afferent information underdetermines the detected property.) The point that has to be understood is that, even with reference to the range of situations to which the system is well adapted, it is better to use an imperfect indicator than to have no access to task–relevant properties. To use veridicality as a filter on this range of situations is to show a touching, but quite unbiological, devotion to truth.

It is misperceptions of this sort which I shall entitle *normal*. They result from the use of indicators that are imperfect but the best available, where this use confers an over-all advantage upon the organism, despite the occasional occurrence of error. Such misperceptions cannot easily be accommodated by a causal theory, because these are cases where, for example, a perception of red is caused by a blue surface even in normal circumstances.¹³

going from nearly white at one end to red at the other. If Land is right, this cloth should not look white at one end, red at the other. It should look uniformly white, and red-lit at one end, and white-lit at the other. Now, take a color photograph of the cloth and print the photograph on the cloth. The cloth will now really be red at one end, shading gradually to white at the other. Now, replace the red lamp by a white lamp of the same intensity. The cloth will look uniformly white, and as before, variably illuminated.

¹³ Normal error is the subject also of Matthen and Edwin Levy's "Teleology, Error, and the Human Immune System," this JOURNAL, LXXXI, 7 (July 1984): 351–372. The concept also emerges clearly in John Haugeland, "Semantic Engines: An Introduction to Mind Design," in his *Mind Design* (Cambridge, MA: MIT Press, 1982), pp. 1–34 (see especially 17/8); and in Christopher Cherniak, "Computational Complexity and the Universal Acceptance of Logic," this JOURNAL, LXXXI, 12 (December 1984): 739–758.

IV. BIOLOGICAL FUNCTIONS: WRIGHT'S ACCOUNT

The first step toward outlining my theory is to introduce the notion of a biological adaptation. Functional explanation in biology seeks to explain why an organism has a particular characteristic (or disposition) by appeal to other characteristics it has as a consequence. There is a good example in the "Science and the Citizen" section of the January 1982 issue of Scientific American. There, under the heading "Thin Red Line," we find a report of speculation concerning the function of the adherence of red cells to immune complexes in mammals. According to this report, the adherence phenomenon had been well known, but thought to have no "functional significance." But it had recently been discovered to be modulated by time and circumstance; it is depressed in human cancer patients, and enhanced in humans with autoimmune diseases. These observations led to two hypotheses, due to different teams of investigators: one team claims that "red cells must be the main agents for clearing immune complexes from the circulation," and the other that they "serve as intermediaries that promote the interaction of a T-cell and its specific antigen, thereby enhancing T-cell activity." The question asked here is of the form: Why does the immune system have complexes that do F? And the answer takes the form: Because as a result of having such entities, the system is able to do G.

Larry Wright provides us with a good foundation for understanding why explanation by reference to functions is applicable to biological entities. His analysis of the meaning of 'Z is the function of X' runs as follows:

- (a) X is there because it does Z, and
- (b) Z is a consequence (or result) of X being there (161).

In the case of biological entities, Wright tells us to evaluate clause (a) above by reference to natural selection. Roughly the idea is this: if the consequence Z is what was responsible for X's evolving or is what is responsible for X's persisting through continued evolutionary change, then Z is the function of X. Otherwise it is not.

Wright's account is distinguished from others by its introduction of a *historical* element into the analysis of function ascription. It is this historical element that enables him to show how a consequence of X's presence can explain X's presence. Consider, for the sake of contrast, a theory that lacks the historical element. Robert Cum-

¹⁴ "Functions," *Philosophical Review*, LXXXII, 2 (April 1973): 139–168; cf. also John Bigelow and Robert Pargetter, "Functions," this JOURNAL, LXXXIV, 4 (April 1987): 181–196.

ion

eks

00-

ce.

of

the

rn-

xes

on

ifi-

me

nd

ns

ne

ng

ley

nd

on

m-

ult

ıd-

g-

X'

(a)

at

ry

on

is

ce

of

n-

ril

mins¹⁵ tells us that functional explanation regards a function as a complex disposition that is explained by showing how simpler dispositions mesh together to result in the complex disposition. Cummins's theory gives an admirable account of the "forward" explanatory direction from X to its function, but not of the "backward" direction from the function to the characteristics by which it is achieved. Consequently, Cummins counterintuitively denies that there is any explanation in the backward direction. But it is precisely this backward explanatory direction which is accounted for by the historical element in Wright's analysis.

The historical element in Wright's account is important also because, if Cummins is right about functions, then there will be no way to choose between rival function ascriptions. We saw earlier that there were two rival hypotheses concerning the "functional significance" of red blood cells' adhering to immune complexes—that it contributes to the removal of immunological detritus, and that it enhances T–cell activity. Now suppose, as is possible, that in fact the adhesion phenomenon does both these things. Then Cummins is unable to choose between the rival hypotheses, because both show how the phenomenon contributes to some disposition of the immune system. Yet, this is a common phenomenon: many features of an organ are such as to have multiple effects, only some of which are functionally significant. What is to prevent us from giving a Cummins—style functional analysis of all or any of these effects?

This consequence of Cummins's view is particularly counterintuitive when it involves effects that actually reduce fitness. The narwhal has a large tusk which actually reduces its mobility. Presumably the tusk is there in order to increase the conspicuousness of a narwhal in mating situations. Obviously, then, the function of the tusk is to increase the sexual attractiveness of the animal; certainly its function cannot be to reduce mobility. Yet, a Cummins—style functional analysis can be given of how the tusk reduces the narwhal's mobility. Wright's analysis is superior to Cummins's because it is able to discriminate between effects on the basis of which one was historically responsible for the phenomenon that is to be functionally explained.

V. BIOLOGICAL ADAPTATIONS

Although I view Wright's analysis as basically correct, it needs to be extended. It is not only the result of an activity that plays a historical role in selection, but also the surrounding circumstances that make the selected activity a good way, or the only way, to achieve the result.

¹⁵ "Functional Analysis," this JOURNAL, LXXII, 22 (Nov. 20, 1975): 741–765.

Digitized by Arya Samaj Foundation Chennai and eGangotri

Consequently, functional explanations found in biology make essential use of facts other than those mentioned in Wright's account.

A good example is to be found in a recent account of ribosome synthesis in the bacterium $E.\ coli.^{16}$ According to this account, the bacterium needs to detect and control the rates at which it synthesizes ribosomes, and the proteins out of which ribosomes are built up. The author reasons as follows: any bacterium acts so as to maximize cell growth and cell division. Now, in order to maximize cell growth and division, the cell has to make enzymes. But enzymes can be manufactured only by ribosomes, so the bacterium must manufacture more ribosomes. On the other hand, the ribosome and its components are energetically expensive to manufacture, and making a surplus is wasteful. Therefore, the bacterium must detect and control the rate of ribosome synthesis and ribosomal protein synthesis, making sure that it is adequate but not excessive.

This exposition endows the activities of the organism not only with functions, but with functional structure. It makes implicit use of functional statements such as:

The function of enzyme synthesis is to facilitate increased metabolism and reproduction.

The function of ribosome synthesis is to facilitate enzyme synthesis.

But it also alludes to certain facts that make some activity *necessary* for the optimal performance of some function. These facts may concern the internal organization of $E.\ coli$ or its parts:

Ribosomes are sites for enzyme synthesis.

Or they may concern $E.\ coli's$ environment (in a broad sense of that term):

Ribosomes are energetically expensive to manufacture.

The role these facts play in the explanation is to subordinate one function to another. The function of ribosome synthesis is subordinated to the function of enzyme synthesis, because the latter is necessary for the former given that ribosomes are the sites at which enzymes are manufactured. The function of controlling the rate of ribosome synthesis is subordinated to the proper performance of 'proper', I mean "such as not to detract from the performance of other functions." Clearly, some measure is required by which to

¹⁶ Masayasu Namura, "The Control of Ribosome Synthesis," Scientific American, CCL, 1 (January 1984): 102–114.

balance off one performance against another: this measure is fitness. In short, facts are used to subordinate one function to another as means to end.

How are we to account for the explanatory value of these facts? The most natural way is to extend Wright's analysis of functions so as to allow that certain facts might have been historically influential in shaping the present constitution of an individual. Thus, consider two activities F and G of an organism such that

(1) G has the consequence F in circumstances C.

en-

me

the

he-

uilt

xi-

cell

can

its

on-

sis,

ith

of

ry

ay

at

ne

li-

S-

ch

of

of

3y

of

to

ri-

t.

(2) G is there because it does F in circumstances C.

Under these conditions, we shall say that G has function F in circumstances C; or, conversely, that G is the realization of F in C. The distinction between activities (G), the functions they serve (F), and the circumstances in which G is subordinate to F, will turn out to be crucial to our account of perceptual presentation.

VI. THE ORIGINS OF PERCEPTION

The above account of functions is adequate for understanding claims about standing features of organisms, e.g., "The function of ribosome synthesis is to facilitate increased metabolism and reproduction." But there is an application of functional explanation which we have yet to touch upon. To see why, consider that biological entities such as fingernails and eyebrows have functions. But fingernails do not, in order to perform their functions, modify their activities to suit changing circumstances—they just sit there.

By contrast, *E. coli's* control of its ribosome synthesis is a much more complex activity: it involves responding differently to different situations. When ribosomes are being manufactured faster than they can be used, there will be idle ribosomes present. These attach to a gene (an *operon*), which then shuts down further ribosome synthesis. When all available ribosomes are involved in enzyme manufacturing activity, the operon becomes unattached again, and ribosome synthesis starts up again. A similar mechanism controls the manufacture of the ribosomal components. Thus, the bacterium's behavior is *plastic*.

It is obvious that the plastic behavior of *E. coli* is subordinate to its functions in exactly the same way as some of its functions are subordinate to others. The control *system* is a permanent feature of the bacterium, and its presence is explained by the need to make neither too many nor too few ribosomes. Given this explanation of the existence of the control system, it seems possible to extend the chain of means and ends down into the plastic activities involved in ribosome synthesis. Let us distinguish between two levels of description. A system can be described either as executing a constant function or as

behaving in a variable manner. These two levels are linked by the fact that, in different sets of circumstances, the system has to perform different activities in order to be executing the constant function. Thus, we have the familiar pattern of variable activity being subordinated to a function, and being variable simply because the facts that mediate the subordination are themselves variable.¹⁷

Now, it is obvious that Wright's historical stratagem will be of no use in explaining this pattern of subordination, since functionally subordinate activities and subordinating circumstances covary concurrently. What is required here is not historical causation, but direct causation. There are two patterns of direct causation which could be used. The system could be taken to be functionally so organized that changing circumstances directly influence functionally directed behavior. Or it could be taken to be functionally so organized as to have internal states which register changing circumstances and which then initiate the activity functionally appropriate to the circumstances thus registered. Broadly speaking, the latter will be used when there is a possibility that behavior appropriate to circumstance C will be initiated even when C does not obtain—the resulting inappropriate behavior may then be explained by the system's falsely registering C. I shall suppose that such internal states are the evolutionary precursors of perceptual states, and ask what would be a suitable schema for attributing content to them. I do not mean to imply that such content attribution is legitimate. My interest in pursuing a content-attribution schema for such states lies solely in the extension of the schema to perception. 18

We have seen that our bacterium *E. coli* does indeed have an internal state that mediates functionally appropriate activity in changing circumstances. When there are idle ribosomes they attach to an operon. When in this state, the operon "switches off" ribosome synthesis. When the ribosome is needed, it gets detached from

¹⁷ Some philosophers have treated variable activity, under the rubric of "goal-directed activity," quite separately from functions—the tendency is to give an account of these in terms of concurrent causation only. See Matthen and Levy (op. cit., note that it eludes these criticisms, which are also dependent on the notion of normal error.

¹⁸ In his "Intention and Action among the Macromolecules," in N. Rescher, ed., Current Issues in Teleology (Lanham, MD: University Press of America, 1986), pp. 65–76, Alexander Rosenberg shows that biochemists have a tendency to describe such states in intentional terms, quoting from a biochemistry textbook which, in the space of a single paragraph, describes tRNA as recognizing, discriminating, undertaking tasks, making errors, editing, correcting, avoiding, and proofreading (65/6). macromolecules" and perception, intention, and action in humans is enough to justify or even compel the use of intentionality. I do not agree with this.

the operon. In this new state, the operon switches on ribosome synthesis. Let us entitle these states of the operon quasi-perceptual states. The question is: What is their content? Intuitively, the right answer is given by the fact that the functional role of the states is to initiate the activity appropriate to the circumstances. The control system has to "sense" when ribosome synthesis is going too fast or too slow, and these are the states which do it. So the ribosome-attached state of the operon is a sensing of excess ribosomes, and the ribosome-free state a sensing of the need for more ribosomes. That is the only answer that links the subordinate behavior to its function in the right way.

fact

orm

ion.

rdi-

that

fno

ially

con-

: di-

ıich

SO

ion-

so so

um-

iate

will

to:

the

sys-

ates

hat

not

rest

y in

an

in

ach

bo-

om

-diunt

ote

unt

mal

ed.,

pp.

ibe the

ler-

/6).

the

This is just where the causal theory has found a foothold; for these states are *caused*, respectively, by excess (i.e., idle) ribosomes and by there being no excess ribosomes. But we have seen that there is reason to reject the causal account, as applied to real perceptual states, because it is incapable of accommodating normal error. Another difficulty for the causal theory is that causes can be redescribed. The states in question can be said to be caused not just by an excess or deficiency of ribosomes, but by the presence or absence of *idle* ribosomes or of molecules with a certain chemical composition, and so on. The causal route does not, in other words, give us a schema that is potentially able to handle the intensionality of perceptual content.

There is another way of getting to what we have identified as the "right answer." Why not use the functional hierarchy to identify the mode of presentation? Might it not be because the function of the state is to trigger the action appropriate to an excess of ribosomes, that we want to say that its content is precisely the presence of such an excess? This way of identifying content is going to yield results quite close to the causal theory, simply because one should expect that, if a state has the function of triggering the appropriate response to the presence of Xs, then it will be brought into being by X_{s} —otherwise, there would be no reliable matching of response and situation, and natural selection would have to "search for" some other way of achieving the appropriate response. In fact, the two theories will diverge precisely in those cases where there is misperception. For, if a state that has the function of bringing about the appropriate response to the presence of xs does come about in the absence of xs, the functional theory would lead us to say that the organism falsely sensed the presence of xs. But this is precisely what we ought to be led to say, is it not? After all, we have given these internal states a role in the theory precisely in order to explain the system's responding to the presence of xs, even when xs are absent.

Digitized by Arya Samaj Foundation Chennai and eGangotri These reflections suggest the following schema for ascribing a "mode of presentation" to the quasi-perceptual states we are considering:

Q: A quasi-perceptual state is a state that has the function of bringing about the functionally appropriate response to the presence of things of a certain sort, where things of that sort are sometimes present. sometimes absent, and

A quasi-perceptual state with internal characteristic S is a presentation of property F if and only if the function of S-states is to bring about the functionally appropriate response to the presence of things that are F.

Note that Q is an inside-out schema for identifying the mode of presentation and that, although it makes the assignment of content to a type of state presuppose an external world, it does not demand that the assignment of content to an individual state imply the existence of external objects. In this way, Q partakes of a weak form of what Putnam has called "methodological solipsism."

VII. TOWARD PERCEPTION

Schema Q has one very obvious disadvantage when it is applied to real perception such as that which we humans experience. Quite simply, there is no such thing as the proper response, or even a range of functionally appropriate responses, to what perception tells us. Suppose I see my glass empty. I may not even come to believe that my glass is empty, since it is possible that my perception is misleading. And, even if I do come to believe that my glass is empty, what I shall do in response to this depends very complexly on my principles, my weaknesses, my desires, and so on, and on my beliefs concerning the state of the world. What time is it? How many have I already had? Obviously, then, schema Q is not going to work: there is no functionally appropriate response mediated by perception.

The most obvious way to retain the intuitions behind the functional schema is to get rid of the reference to the proper behavioral output of perception. Suppose we think of perception not as leading directly to action, but as leading to epistemic and practical assessment, which, in turn, leads to action. Can we not then define the output of perception in terms of its input to these faculties? The following is an attempt to do this:

P: A perceptual state is a state that has the function of detecting the presence of things of a certain type, where things of that type are sometimes present, sometimes absent, and

A perceptual state with internal characteristic S is a presentation of property F if and only if its function is to detect the presence of a thing This schema is open to an objection that we must now attempt to overcome. The objection is based on the view that epistemic and practical assessment feed into and modify perception as much as perception feeds into assessment. Nelson Goodman puts the point

with characteristic panache:

ga

on-

of

ent nd

he

rm

to

ite

ge

IS.

ny

ıg.

all

ny

he

d?

C-

C-

al

ıg

S-

ne

ie

The Kantian dictum echoes here: the innocent eye is blind and the virgin mind empty. Moreover, what has been received [i.e., by perception] and what has been done to it [by reason] cannot be distinguished within the final product [belief]. Content cannot be extracted by peeling off layers of comment.¹⁹

If Goodman is right, it is not possible to speak of perception as detecting anything. We have beliefs about what is the case and no other epistemically significant states that constitute, as it were, prebeliefs. So to speak of the *content* of what perception contributes to belief and action is an impermissible abstraction.

It may very well be right to say, as Goodman does, that perceptual content cannot be got at by peeling off layers of judgmental content. But I think that it is possible, nevertheless, to get at what perception contributes to judgment, simply because it is possible to create situations in which the output of perception gets isolated by the fact that judgment is not allowed sufficient information with which to comment on it. Techniques for achieving this result experimentally are in evidence in the pioneering work by Bela Julesz and Land.²⁰

Julesz was dealing with depth perception by means of measurements of binocular disparity, i.e., the displacement of an image of something on one retina with respect to the other. The big problem that has to be solved before this sort of depth perception can fully be elucidated is this: How do we manage to identify two images as of the same object—how do we match images in order to measure their binocular disparity? Some psychologists of perception have tried to solve this problem at a very high level; they proposed that the world is first segregated into big external objects (such as geometrical figures on a ground or even objects such as pieces of furniture or human beings) by means of some innate (so says the Gestalt theory) or learned ("New Look" theories) process of identification, and that the images of these objects are then matched. This is the type of proposal which leads to Goodman's sort of skepticism about the separability of perception and other mental processes, for it makes

¹⁹ Languages of Art, 2nd ed. (Indianapolis, IN: Hackett, 1975), p. 8.

Julesz, "Binocular Depth Perception of Computer Generated Patterns," Bell Systems Technical Journal, XXXIX (1960): 1125–1162; and Land and McCann, "Lightness and Retinex Theory," op. cit.

judgment influence perception just as much as perception influences judgment. If this sort of theory is right, it would indeed be wrong to identify the output of perception as the input to judgment.

Julesz disproved the contention that the measurement of binocular disparity is a process that operates at a "high level" by recreating depth perception on "random—dot stereograms." These stereograms consist of two images: one is a computer—generated random pattern of dots, and the other is a repetition of the same with a square region shifted over by a few minutes of arc. When these images are presented, one to one eye and the other to the other, it appears that the square region is closer to the observer than the rest. This shows that the binocular disparity of the square region was detectable. But, because the pattern consists only of randomly placed dots, the measurement could not have needed prior segregation of objects within the image.

Land's experiments concerned the perception of the brightness of an object. (When the brightness perception from all color channels is combined, this sort of perception amounts to color perception.) We have seen that the perceived brightness of an object exhibits a kind of constancy not displayed by the light it reflects under varied circumstances—for example, a uniformly colored wall will look uniformly colored, even when it is half in shadow. Now, this is a natural place for a high-level theory. How are we to explain the fact that different light information culminates in the same appearance? Why not say that our belief that the wall is uniformly colored informs our perception of it as uniformly colored? Land disproved this supposition by showing that humans are able to detect color constancy even in random patchworks of color-"Mondrians" as he called them, because of their resemblance to the abstract works of that artist. The Mondrians are presented to observers who have never seen them before; and these observers are able to discriminate different colors that are so illuminated as to send the same signal to the eye, and identify patches of the same color even when they are so illuminated as to send different signals to the eye. The constancy phenomena are thus displayed in the complete absence of any beliefs about the actual brightnesses present. Moreover, the Mondrians lack objects with which colors or brightnesses are inductively associated—apples, leaves, etc. So another source of interference by beliefs is eliminated.

I conclude from these results that it is at least sometimes possible to isolate what is detected by perception independently of judgment, decision, and the like. Whether or not it is true that perception can be influenced by judgment, this suggests that what perception detects exists independently of judgment. We shall now pursue schema

P, the thesis that a perceptual state presents that which it is its function to detect.²¹

VIII. REALISM AND THE FUNCTIONAL SCHEMA

It is a consequence of schema P (the functional schema, as I shall call it) that the properties presented by perception exist independently of perception. Understood in the light of the functional schema, Land's theory is a hypothesis about the functional organization of color vision. The function of color vision is to detect surface color. The (supposed) activity of the visual system in detecting brightness "edges" in the image incident upon the retina and in processing this information in the way described earlier is subordinate to the function of detecting color. The fact that mediates this subordination is, simply, the correlation between brightness edges and color discontinuities. It was, in other words, the historical operation of the correlation, together with the advantage of possessing a color detector, which was responsible for the evolutionary persistence of this edge processing.

n

a

S

f

f

The claim that content can be attributed to perceptual states as a result of such hypotheses about functional organization has two immediate consequences. First, it provides a justification for expressing the content of color vision in distal terms. Suppose that somebody were to say that Land's theory shows not that we see color, but that what we have so far mistakenly supposed to be color vision is, in fact, a system for detecting different sorts of brightness gradients in the (proximal) image incident upon the retina. This hypothesis about content is perfectly compatible with everything that Land says about the mechanism of color vision as it now exists. Only when we take the evolutionary development of color vision into account do the two ways of describing Land's theory turn out not to be factually equivalent. The functional schema gives crucial importance to the (supposed) fact that the ability to distinguish different sorts of brightness gradient in an image does not by itself confer an evolutionary advantage upon us. Rather, this discriminatory ability persists only because, as it happens, it happens to correlate closely with the ability to discriminate surface color.

A second important consequence of functionally identifying the

²¹ In *The Modularity of Mind* (Cambridge, MA: MIT Press, 1983), Fodor argues that in fact judgment has a very limited influence on perception: there is no feedback to perception except when calls on memory are built into the computational routines that constitute perception. Similarly, Marr allows memory to influence perception only after what he calls the "2 ½ D sketch" has been formed. Both these conditions constitute very significant restrictions on the unlimited feedback that Goodman and others of his ilk posit. But I have tried to make my argument independent of these restrictions.

output of the system we are discussing as information about *color* is that it makes the system subject to what I earlier entitled normal misperception. As we saw in section III (and note 12), it is possible to create a surface on which color varies spatially in the manner normally characteristic of illumination gradients. Such a surface would, when uniformly illuminated, be perceived as uniformly colored. The functional schema enables us to identify this as a misperception—which is surely what it is.

Together, distality and the possibility of misperception constitute realism in the theory of perception. Distality implies that perception purports to be of features of the world which are independent of perception itself. That, of course, does not imply realism, since it does not imply the actual existence of the independent world. But add the possibility of misperception, and we now have realism; for misperception would not be possible if the world of perception did not transcend perception itself.

In being thus committed to realism in my interpretation of the new theories of perception, I find myself at odds with a number of philosophers—Keith Campbell, C. W. K. Mundle, Clyde Hardin, and James McGilvray—who have discussed computational theories of perception in general, and color vision in particular. They have two lines of argument in support of an antirealist interpretation.

The first line of argument, which is emphasized by Hardin and Campbell, ²² assumes that a color, blue for example, will be defined as the real property that, in normal circumstances, produces a particular sort of quale in us. But, they say, there is no such property.

Campbell and Hardin give two sorts of examples in support of this view. The first kind is irrelevant. Hardin says: "Sapphire is blue because of the transference of electrons from iron to titanium, but lapis lazuli is blue because of the vibrational energy characteristics of conjugated bonds" (496; cf. Campbell, 137). So what? It does not follow that they have no relevant real property in common—for example, reflectance.

The second sort of example is more troubling, at first sight. Perceived color is not all due to reflectance. Some surfaces may be blue because of reflectance, but the sky is blue because of dispersion. There is no real property common to such diverse causes of color perception. I shall make two comments about this. In order to make the point that Hardin and Campbell want to make, one must show that misperception is not involved. Why should we not say that the

²² Campbell, "Colours," in Robert Brown and C. D. Rollins, eds., *Comtemporary Philosophy in Australia* (New York: Humanities Press, 1969), pp. 133–157; and 491–500.

is

al

to

rd,

ne

te

n

of

it

ıt

r

d

W

S-

f

d

sky is misperceived as blue? Maybe the evolutionary advantages of color vision pertain only to the discriminatory resources it affords with respect to surfaces. Perhaps perceiving the sky as blue is just an artifact of this mechanism and it neither increases nor decreases the fitness we get as a result. That would be the sort of case in which we should want to say that the sky is (normally) misperceived as blue. Thus, the fact that it is perceived as blue in normal circumstances does not imply that it is blue.

There is, however, a more general point in Campbell and Hardin's arguments which needs to be considered. We perceive not only reflecting surfaces as colored, but also translucent, irridescent, and luminous surfaces. The reflectances of these other surfaces is quite irrelevant to their perceived colors. Surely it makes no sense to say that these are all misperceived. But this still does not prove the conclusion. What it proves is that it is not *reflectance* that we perceived, but some extension of reflectance, some property that is coextensive with reflectance when restricted to reflecting surfaces. I shall not attempt to define this property here, but I believe that it would not be hard to define.

A second line of argument is emphasized by Mundle and McGilvray.²³ These philosophers argue that, according to Land, our awareness of what we call "color" is the product not only of the stimulations incident upon the retina, but of our own "mental" activity. Thus, they assimilate color vision to something like the Kantian model of transcendental idealism, arguing that, though color vision is initiated by a signal from without, the end product represents the external world in a way that owes more to the operation of color vision than to the nature of properties that have real existence.

The position taken by these philosophers is, I think, essentially the same as supposing that "color" vision is actually a presentation of certain kinds of brightness gradient. The claim is that there is no way to identify the output of color vision except as the input stimulation processed in a certain way. They do not question that this output should be *called* "color": their point is that, whatever it is called, there is no reason to identify that output as representing an external property. I claim, of course, that this is wrong: the functional schema—and the historical explanations it implicitly invokes—provides us with just this reason.

IX. INTENTIONAL EXPLANATION

A strength of the functional schema is that it accounts well for the peculiarities of intentional explanation—the explanation of the

Mundle, Perception: Facts and Theories (New York: Oxford, 1971): chs. 9 and
 11; and McGilvray, "To Color," Synthese, LIV, 1 (January 1983): 37–70.
 CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

causal interrelations of states by reference to the content assigned to them.

The first point to make here is that the functional schema enables us to distinguish between states that should be assigned content from states that should not. In this respect, the causal theory is not helpful. Recall that, according to a causal theory, the content of a perceptual state is what normally causes states of that material type. There is nothing in this theory itself which prevents us from saying that, for example, a column of smoke is a perception of wind speed, since it has a feature—the angle it makes with the horizon—that is normally caused by a particular wind speed.24 Thus, a causal theory needs an adjunct that distinguishes content-bearing states from non-content-bearing states. This does not apply to the functional schemata, Q and P. These assign content only to states that have the function of registering the presence of some object or circumstance. Since the angles of smoke columns do not have functions, they are excluded from the purview of this schema.

The second point to make is that the functional schema helps us understand the parallel relationships employed in intentional explanation. In intentional explanation, the explanantia and explananda are states with content. The explanation presupposes that the explanantia states cause the explanandum state to come into being²⁵; the explanatory weight, however, lies not on this causal relationship, but on the logical relationship that exists between the intentional objects of the states. For example, suppose we seek to explain X's belief that p in terms of his belief that q and his belief that if q then p. This explanation would be inadmissible if the explanantia beliefs had not caused the explanandum belief to come about. The force of the explanation, however, is not primarily due to this causal relationship, but is due to the logical relationship that holds between q, if q then p, and p. Once we are told the relationship of the content of \hat{X} 's belief to the content of his other beliefs, we understand why he believes that p. It is this obliqueness of accounting for states by reference to content which characterizes intentional explanation.

It is puzzling that it should be legitimate, in explaining why X came to believe that p, to appeal to the logical relations that p bears to the objects of X's other beliefs—on the face of it, this explanatory strategy involves a category mistake, a confusion between the propositions toward which beliefs are directed, and the beliefs themselves.

²⁴ The example is David Hills'; see "Mental Representations and Languages of Thought," in Ned Block, ed., Readings in the Philosophy of Psychology, vol. 2 (Cambridge, MA: Harvard, 1981), pp. 11-20.

That intentional explanations require causation is, of course, a point rediscovered by Donald Davidson in "Actions, Reasons, and Causes," this JOURNAL, LX, 23

For, if we are interested in the transition from beliefs already held to new beliefs, we ought to attend to the causal powers of beliefs. Why should these causal powers be determined by abstract semantic relations in the realm of propositions? To answer this question is essentially to show why intentional explanation is legitimate. And, if a particular scheme of intentional ascription makes it easy to answer the question, that increases its plausibility as well as the plausibility of the explanations it supports.

It is easy to see that the functional schema helps explain why states should be connected in this way. If the function of a state is to detect the truth of p, and of another state to detect the truth of if p then q, then, if the detection of q is of any advantage at all, we should expect that a connection might well develop between these states and the state that detects q. By contrast, a causal theory is of no help here at all. That a state is normally caused by the truth of p, and another one by the truth of if p then q, is no reason at all to expect that the two together would cause the state that is normally caused by q. This is a consideration which applies not just to perceptual states but to all states with content—it suggests that they should all be defined functionally.

Evolutionary epistemology has sometimes been regarded as providing us with reasons for disavowing skepticism. It is easy to see now that it cannot do this. For evolution may provide us with senses that are quite reliable, but there is no reason to believe that it provides us with senses that guarantee the truth of any particular perception that is the lesson of "normal misperception." Nevertheless, evolutionary epistemology has an important role to play. We started this paper with a question from traditional epistemology-What constitutes the mode of presentation in perception? We found that computational theories of perception have something important to contribute to this question and that the theory of evolution is foundationally important to computational theories. To the extent that such theories tend to inform us about the structure of our perceptually derived concepts, then, evolution can after all do what Konrad Lorenz,26 the founder of evolutionary epistemology, hoped. It can show how that which is a priori for the individual is a posteriori for the species. That is, a study of evolution can help us lay the foundations of an understanding of innate ways of representing the external world. That is no small contribution.

MOHAN MATTHEN

University of Alberta

²⁶ See "Kant's Theory of the *A Priori* in the Light of Evolutionary Biology," in Konrad Lorenz₂₀Rin bulkarosmetanesu (New Argric dilateon partial 93:5), pp. 181–217.

BOOK REVIEWS

Having Reasons: An Essay on Rationality and Sociality. FREDERIC SCHICK. Princeton, N.J.: University Press, 1984. 159 p. Cloth \$22.50, paper \$7.95.

Herbert Simon has coined the term 'substantive rationality' to denote the pattern of consistency of the behavior that a sane person typically exhibits. Explaining what it means for behavior to be consistent in this way is the aim of "decision theory," a family of normative and positive theories that characterize rational action in terms of the agent's beliefs and desires. Beyond this general form of explanation, these theories tend to have in common the assertions that a rational agent chooses actions on the basis of their consequences, that these consequences are external to the agent, and that the agent's desires can be represented by a binary preference relation between consequences. An agent is substantively rational if his preference relation possesses some formal characteristics (e.g., transitivity) and if his pattern of action is consistent with achieving consequences to which no other achievable consequence would be preferred.1 These ideas lead to an elegant, parsimonious theory which has been useful in economics, statistics, and other social sciences.

Because the standard versions of decision theory do not systematically distinguish between self-regarding and other-regarding preferences, it has sometimes been argued (a) that the theory is intrinsically concerned with selfish preferences only, and (b) that a different theory of action is needed to account for the rationality of actions motivated by concern for others. In Having Reasons, Frederic Schick considers both of these arguments, rejecting the former and endorsing the latter.

Schick begins (chapter 2) by giving an informal exposition of a version of decision theory in which the agent's beliefs are represented by probability intervals. For instance, the agent might believe that the probability of rain is between 10 and 60 per cent, without having any more specific belief. The agent might prefer to carry an umbrella if he ascribed 60 per cent probability to rain, and not to carry one if he ascribed 10 per cent probability. So his vague belief gives him no ground for having a preference one way or another. Schick stresses that this lack of preference is a different phenomenon

¹ For brevity, I ignore here the complication of actions having uncertain consequences.

from indifference that occurs when the agent makes an affirmative judgment that the alternatives are equally good, and he argues that the incompleteness of preference can allow the theory to avoid counterintuitive implications of the standard theory which have been pointed out by Maurice Allais and by Daniel Ellsberg.

IC

th

e-

n-

a-

ρf

a-

a

Schick suggests (chapter 3) that utility maximization should be viewed as being only the first among several criteria of choice. When this criterion does not definitely recommend any action (as in the umbrella example above), further criteria are considered. One of these criteria is a causal version of decision theory, which is distinct from the epistemic version developed in chapter 2.3 Schick mentions that Newcomb's paradox is a situation in which the preferences determined by epistemic utility alone would plausibly be incomplete, because the agent would not have a unique, tightly specified probability judgment about the behavior of Newcomb's demon. He argues that the agent would thus appeal to causal decision theory to choose an action, so that his theory resolves this paradox in the same way as does the causal theory, rather than in the way that most epistemic versions of the theory resolve it.

Next (chapter 4) Schick uses this theory as a basis for a theory of other-regarding, or *responsive* preferences. He applies this theory to account for the possibility of cooperation in situations that, if each agent were to act solely on the basis of a narrow conception of his own welfare, would have the prisoner's-dilemma characteristic that a unanimously preferred outcome is foregone. Schick assumes that the agent's beliefs are ambiguous, as in his discussion of Newcomb's paradox; so again the agent appeals to causal decision theory to make a decision. Thus, the discussion of cooperation could have been based on the causal version of the theory rather than on Schick's own version. It will be helpful briefly to consider the discussion from this perspective.

It is possible for cooperation among selfish agents to be rational if their utilities are evaluated unambiguously (contrary to Schick's supposition) according to an epistemic notion of conditional probability. The fact that a noncausal version of decision theory can explain cooperation in this way without positing responsive preferences does not make Schick's explanation less interesting, though. In fact, to an adherent of the view that causal decision theory is the appropriate version of the theory for characterizing the substantive

³ A decision theory is called *epistemic* if expected utility is defined from conditional probabilities, and *causal* if it is defined from probabilities of counterfactuals.

² Schick actually adopts a terminology according to which actions determined by this criterion alone do not involve choice.

rationality of actions, there are two reasons to welcome Schick's explanation. The first is simply that cooperation is thought to be rational. The second is that the absence of a causal account of rational cooperation had seemed to be a consideration against adoption of the causal theory, and now that consideration is removed.

One troublesome matter is the analogy that Schick draws between an agent's responsive preferences and Kenneth Arrow's concept of a social-welfare function (103). The point of Arrow's work is that a social-welfare function satisfying a few apparently reasonable conditions must be dictatorial. This conclusion does not leave room for other-regarding preferences (except that possibly an agent could have completely selfless devotion to one other person), and Schick does not indicate in what respect the analogy is inexact so that the conclusion may be avoided.

Finally (chapter 5), Schick introduces another secondary criterion, besides causal decision theory, that an agent can use when epistemic utility comparisons are not clear cut. This criterion is *sociality*, which involves the direct effect of another person's interests on the agent. Schick points out the resemblance in some respects of this concept to others, notably Thomas Nagel's concept of altruism.

Until this point, Schick's theory has invoked the same concepts as have earlier decision theories. The thrust of the argument has been that, by applying an epistemic and a causal version of the theory lexicographically (that is, by using the causal version to break ties), several cases that are problems for other versions of decision theory can be resolved. Sociality, though, is not a familiar concept from decision theory. In particular, the concept of sociality is not reducible to the agent's beliefs and desires. In Schick's terminology, the introduction of sociality goes beyond rational-choice theory and constitutes social theory. The rest of this review will be devoted to examining Schick's arguments for the necessity or desirability of such a theory.

Since Schick has already succeeded in explaining cooperative behavior in terms of rational-choice theory with responsive preferences, the need to explain such behavior cannot be the sole justification for introducing sociality. Rather, Schick takes the position that a family of explanations involving notions such as social bonding and personal commitment are also legitimate, and he argues (chapters 5, 6) for the plausibility of reducing these notions to a concept of sociality.

There are two ways that Schick's position might be read. One is that there are facts (Schick discusses Stanley Milgram's experiments regarding obedience to authority and Richard Titmuss's study of 's

n

a

a

1-

r

d

k

e

n

blood donation) that are "hard cases" (7) for rational-choice theory. The other is that rational-choice theory can account for the facts, but that an alternative explanation based on sociality is nevertheless more appropriate. Indeed, it is not easy to provide a rational-choice-theoretic explanation of Schick's hard cases, but difficulty is not tantamount to impossibility. If the notions of commitment and social bonding can be given a rational-choice-theoretic interpretation or reduction, then Schick's position would have to be rejected if the former reading were intended.

A rational-choice-theoretic interpretation of commitment does seem feasible. For example, imagine Bach's interview by a wealthy nobleman. The nobleman tells Bach that in principle he would like to hire him as director of court musicians, a position which would be much more remunerative than Bach's current one but which also would involve a move to a rather unattractive town. The nobleman also expresses the concern that Bach would quickly decide to move back to beautiful Leipzig, where his current patron would be happy to reinstate him. To allay this concern, Bach and Mrs. Bach have their seventeenth child, enlarging their household to a size that is absolutely impossible to maintain on Bach's current salary. Their choice to have this child is a commitment to make a permanent move. This choice can be explained in terms of the Bachs' desires and beliefs, as a renunciation of their most preferred conceivable situation (having a high salary and sixteen children) in order to obtain their second favorite (high salary and seventeen children) rather than being left without any alternative to their least preferred situation (low salary in Leipzig and sixteen children). This renunciation, when the only real choice is between the second and third best alternatives, is evidently recommended by rational-choice theory.4

The objection might be raised that Bach's behavior in the preceding example is self-interested: Schick is trying to explain other-regarding commitments, but the preferences described in the example are self-regarding. A parallel example could be based on responsive preferences as well. Thus, the rational-choice-theoretic reduction of commitment should be able to explain the facts that Schick presents; so the objection should be rejected. If social bonding can be given a rational-choice-theoretic interpretation similar to that for commit-

⁴ The example just described bears close resemblance to Newcomb's paradox, with the nobleman playing the role of Newcomb's predictor and with having an additional child corresponding to the action of taking only one box. However, it is envisioned here that the nobleman can offer employment after the child is born, whereas it is essential to Newcomb's paradox that the predictor has already acted irrevocably before the choice of boxes is offered.

ment (and Schick has not provided any reason to doubt this), then the "hard cases" for rational-choice theory are merely hard—not

impossible to explain.

The other possible reading of Schick's justification for introducing sociality: that an explanation of commitment and social bonding in terms of sociality is somehow superior to rational-choice-theoretic reduction of those concepts, imputes to Schick a position that needs to be fleshed out. If there is a criterion according to which the former explanation is superior, it has not been explicitly described. At the beginning of the chapter concerning commitment, Schick does write that "a person who admits a commitment is . . . applying his ethics," (120) and that "To have an ethics is to have ideals" (121). Thus, it might be consistent with his position to assert that commitment has an idealistic character that the prudential choices recommended by rational-choice theory lack. Logically, this position would be an instance of a familiar objection to various attempts at scientific reduction: that one entity cannot be "nothing but" a construct of other entities if it possesses a property that the entities to which it is reduced lack. Other instances of this objection (for example, the objection to statistical mechanics on the grounds that atoms do not have heat) are not widely accepted, and there is no reason to think that Schick accepts them. Thus, beyond its mere logical consistency with what Schick has explicitly written, there is no reason to impute to him the view that commitment and social bonding ought to be characterized in terms of other forms of idealistic behavior or feeling. There is no other likely candidate for a criterion that Schick might be adopting according to which sociality is a superior theory, either, and such a candidate seems to be needed.

The arguments in the preceding five paragraphs are directed toward a conception of the theoretical status of sociality which is stated at the beginning of the book. Schick writes that "The project I have in view is to design a theory" (6) and that "I propose a concept of sociality as a step toward a theory that works better. That theory is then developed" (7). On this conception, reasons ought to be presented that would justify the creation of a new theory of social behavior. It has been argued here that Schick has not succeeded in presenting such reasons. It is possible, however, that these arguments are misdirected, because the passages just quoted poorly represent what Schick is actually trying to do. In particular, although those passages seem to promise that a formal deductive theory will be presented in that way, and Schick suggests that it might not be feasible or even desirable that it be so (102).

Accordingly, it may be misleading to describe the book as a comparison of two theories of choice and action in social situations. Rather, the later chapters of the book could better be characterized as an attempt to clarify the logic of a family of common-sense explanations of social actions and to defend these explanations as being cogent (and useful to social scientists) even though they may not meet the ordinary criteria of scientific explanation. That is, Schick's thesis might best be stated as something like: (a) it is possible that rational-choice theory ought to be the basis for the logic of justification in the social sciences, but (b) concepts like commitment and social bonding play an important part in the logic of discovery, and (c) their role is best understood by interpreting them in terms of a more comprehensive concept of sociality.

EDWARD J. GREEN

Economics, University of Pittsburgh

C

C

The Varieties of Reference. GARETH EVANS. JOHN MCDOWELL, ed. New York: Oxford University Press, 1982. xiii, 418 p. Cloth \$34.95, paper \$10.95.*

This posthumous work, edited by John McDowell (who is to be commended for his excellent job), is not the book that Gareth Evans would have published had he lived. But it is rich, enormously subtle, ingenious, well written, if sometimes obscure, and obviously the product of an absolutely first-rate philosophical mind. The book is about the sense of singular terms and about the thoughts expressed by utterances of sentences containing singular terms. One might expect a work on these topics to be based on a general theory of meaning and content, but this is not Evans's approach, and I think that this lack of a foundational basis may make it difficult for some readers to relate much in the book to their own foundational concerns.

THOUGHT

Most of the book is about the nature of beliefs about ordinary physical objects, the kind of belief that would be expressed by utterances of sentences such as 'This is a necklace', 'I am in the tub', 'That peach was very fuzzy', and 'Mick Jagger has never written an opera'. I shall

0022-362X/88/8501/0033\$01.00 · © 1988 The Journal of Philosophy, Inc.

^{*} I am grateful to Kent Bach, Mike Harnish, and Brian Loar for their comments on an earlier version of this review.

now list, and comment in passing on, some of Evans's central theses about singular thoughts, as he calls the kind of thoughts in question. After this section on thought I shall have a similar but briefer section

on language.

Singular thoughts are "Russellian." Evans uses the expression 'Russellian singular term' to mean "a member of a category of singular terms such that nothing is said by someone who utters a sentence containing such a term unless the term has a referent," and it is "a convenient extension of terminology . . . to speak of thoughts as being Russellian: a thought is Russellian if it is of such a kind that it simply could not exist in the absence of the object or objects which it is about" (71). It may not be clear what exactly he intends by this, but it is at least clear that, for Evans, there is no way of specifying the content of a Russellian thought about an object without referring to that object. This is in contrast to what would be held by a theorist who claimed that a thought was about an object only if it was so "by description"; for this theorist would hold that, when a subject's belief is that a certain object is F, that is so because, first, for some property G, the belief is also a belief that the G is F, and, second, the object in question is uniquely G.

Although Evans recognizes Russellian thoughts, he rejects a further Russellian thesis about the nature of Russellian thoughtsnamely, the "ordered-couple conception of Russellian thoughts," the idea that "the content of a Russellian thought involving the ascription of a monadic property to an object can be appropriately represented as the ordered couple of the object and the property" (81). To know the content of a Russellian thought it is not enough to know what object the thought concerns and what is thought about it; one must also know the mode of identification for the object, the way of thinking of it, employed in the thought—what Frege called a mode of presentation. It is possible "for two thoughts to ascribe the same property to the same object while differing in content, because they differ in the ways in which the object is identified" (83).

Thus, suppose that, looking at Al, I think, "He's drunk." A theorist who holds the ordered-couple conception of Russellian thoughts would represent the content of my thought by the singular proposi-

(Al, drunkenness)

But Evans would reject this and insist that any specification of the content of my thought must include a reference to the way in which I am thinking about Al, the mode of identification—in this case, what Evans calls a demonstrative mode of identification—that my thought CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

employs and by virtue of which my thought is about Al. This suggests that the content of my thought could be represented (at least at a certain level of analysis) by the different ordered couple

ses

on.

lon

on

ξu-

ice

as

t it

1 it

out

he

to

ist

by

oe-

ne

he

ır-

he

ely

to

it;

ne

la

ne

se

0-

ts

si-

ne

I

at

nt

$\langle M_{\rm al}, \, drunkenness \rangle$

where $M_{\rm al}$ is the particular demonstrative mode of identification of Al employed in my thought.

But what are these "modes of identification," these "ways of thinking about things" which enter into the contents of Russellian thoughts? We are certainly entitled to ask this question, because, until the notion is explained, expressions such as 'mode of presentation/identification' and 'way of thinking about a thing' are simply suggestive terms of art for whatever is needed to complete the content of a Russellian thought once one has been given the object the thought is about and what is thought about it, and because one might be skeptical of there being such content-completing things.

Evans recognizes the need to explain what he means by a "way of thinking about an object," and that, in effect, is what most of his book is about. The main constraint on our understanding of modes of identification is that they must conform to *Russell's principle*, which I shall take up presently. At the same time, Evans suggests early in his book that the desired notion of a mode of identification, or way of thinking about an object, "can be explained in terms of the notion of an account of what makes it the case that a subject's thought is a thought about the object in question" (20). Such an account will take the form

[*] S's thought at t concerned x in virtue of the fact that R(S, x, t)

and different accounts of this form, utilizing different replacements for 'R', will yield different ways of thinking about a thing (20, 200).

I am puzzled by Evans's claim that his notion of a mode of identification can be explained in terms of an account taking the form [*] of what makes a subject's thought about the particular object the thought is in fact about. I would have thought that a replacement for 'R' in the sentence form

My present thought is about Al in virtue of the fact that R(myself, Al, the present time)

could explain what made my thought about Al only if it was not in terms of my having further thoughts about Al. But it is impossible to

¹ Evans recognized that "modes of identification" would also be needed for the properties our thoughts concern, and he looked forward to a future occasion when he could deal with this. I shall follow Evans's book in ignoring this issue.

Digitized by Arya Samaj Foundation Chennai and eGangotri

find in Evans ways of replacing 'R' which are not question-begging in precisely this way. Although most of Evans's book is about the non-descriptive modes of identification which are individuative of the contents of Russellian thoughts, his account of the various types of modes of identification—demonstrative identification, self-identification, and recognition-based identification—are always couched in terms of thoughts about things involving precisely the types of modes of identification in question. I hasten to add that there may not be anything wrong with this per se, but it does seem to preclude Evans from explaining the notion of a mode of identification in terms of "an account of what makes it the case that a subject's thought is a thought about the object in question." Presently I shall take issue on another problematic feature of Evans's explication of modes of identification in terms of thoughts involving those very modes of identification.

Russell's principle, which holds that "a subject cannot make a judgement about something unless he knows which object his judgement is about," has a reading on which it is true and important. "The difficulty with Russell's Principle," Evans is quick to acknowledge, "has always been to explain what it means" (89). Evans, no doubt wisely, does not choose to give a careful explication of the principle and then show how the modes of identification employed in our Russellian thoughts enable us to satisfy it. Rather, he leaves us, for the most part, to gather the intended sense of the principle from the various features he ascribes to Russellian thoughts in his discussions that fall under the rubric of Russell's principle. I say "for the most part" because the use of Russell's principle is importantly guided by Evans's understanding of it as requiring "discriminating knowledge: the subject must have a capacity to distinguish the object of his judgement from all other things" (89) and by his understanding that the following would not be a case in which a subject has this capacity. Suppose that a subject sees a steel ball suspended from a point and rotating about it, and that on the next day he sees a different but qualitatively indistinguishable ball rotating about the same point. And let us further suppose that the subject retains no memory of the first episode "because of a localized amnesia produced by a blow on the head" (90), but that he later reminisces about "that shiny ball" he saw (perhaps he thinks, "That shiny ball was pretty"). In this case, Evans holds that, even though just one steel ball was causally responsible for the present memory, the subject would not have discriminating knowledge of it and would not, therefore, have any beliefs about it [Evans takes himself to be entitled to assume that, this being a normal case, the subject "would not think of distinguishing the ball

he is thinking of as the one from which his current memory derives" (90)]. I shall return to this example presently. As a structuring device Russell's principle is very important for Evans, and I should try to convey something more of what he says about thoughts in its name.

Consider again my demonstrative thought that Al is drunk. According to Evans, this thought is structured in that it has something in common with the thought that Al is sober and something in common with the thought that Betty is drunk. The thought, he would say, is a joint exercise of two distinguishable abilities, one, which is my concept of drunkenness, permitting me to have thoughts that ascribe drunkenness to persons other than Al, and the other, which Evans would call my 'idea of Al', permitting me to have thoughts that ascribe to Al properties other than drunkenness. "An Idea of an object, then, is something which makes it possible for a subject to think of an object in a series of indefinitely many thoughts, in each of which he will be thinking of the object in the same way" (104; my italics). Since Russellian thoughts obey Russell's principle, our ideas of objects must afford discriminating knowledge of the objects they are ideas of, and the chapters on the different kinds of modes of identification are intended to show how the different kinds of ideas, such as demonstrative ideas and recognition-based ideas, yield the knowledge required by Russell's principle.

The entire discussion is constrained and guided by an important thesis which is stated in terms of the notion of a fundamental Idea, which is defined in terms of the notion of the fundamental ground of difference of an object at a time. The fundamental ground of difference of x at t is what of necessity differentiates x from all other objects at t. One has "a fundamental Idea of an object if one thinks of it as the possessor of the fundamental ground of difference which it in fact possesses" (107). The important thesis is this: If α is for S a nonfundamental idea-of-an-object (and thus something enabling S to think a proposition of the form $\Gamma \alpha$ is F^{-3}), then α consists in S's "knowledge of what it is for an arbitrary proposition of the form $\Gamma \delta = \alpha^{-1}$ to be true" (110), where δ is a fundamental idea. This is then brought to bear on the above steel-ball example. For let δ * "be a

statue and a piece of clay" (107).

This is Evans's terminology and his use of corner quotes; he warns, however, that talk of an idea occurring in a proposition must be understood as a mere façon de parler.

² What distinguishes a physical object at a given time is, evidently, that, for some individuating *kind* of object to which it belongs, it is the only object of that kind to occupy a particular spatial region at that time. The appeal to a kind is needed because, "although two Gs may not be able to share a position at a time, a G may be able to share a position with a thing of a different kind: for instance a statue and a piece of clay" (107)

fundamental Idea of that object from which the subject's current conception ('that ball') causally derives"; then the reason the subject fails to have a thought that satisfies Russell's principle—i.e., fails to have discriminating knowledge of the relevant steel ball—is that he does not know "what it is for the identification δ = that ball to be true," and this is because "he has not an inkling of the kind of consideration that in fact would *make* it true" (117).

I have several difficulties with this elaboration of Russell's principle. First, Evans's position entails that one cannot have a belief about something unless one knows the sort of fundamental ground of difference it has; but, although this may have some plausibility as regards common physical objects, it does not seem to be a general requirement for singular thought. Surely I may have beliefs about Alfred's arthritis—I may think, "His arthritis must be painful"—but I certainly have no idea of what the fundamental ground of difference for arthritis is. Second, let us grant that the subject in the steel-ball example does not satisfy Russell's principle, when this is understood as Evans intends it to be; Evans must then show that the example is not a counterexample to Russell's principle, and on this matter he is not very persuasive. We are to imagine a subject who has a memory image of a certain steel ball and who thinks, and perhaps says, "That ball was pretty." Most people, I think, would be prepared to say that the subject believes that the ball he is thinking of is pretty. Evans discusses such apparent counterexamples to Russell's principle and points to cases where we can say that someone is thinking of an object even though he does not have any beliefs about it; but I can find nothing in this discussion to show that what most people would be prepared to say about the steel-ball case is not strictly true.⁴ Third, I am bothered by Evans's argument to show that the subject in the steel-ball example would not have discriminating knowledge of the ball from which his current memory conception causally derives. The claim is that the subject lacks discriminating knowledge in that he would not know what it would be for the proposition this ball = that ball to be true, and this because he would not know what would make it true. I take it that the subject would not know what would

⁴ Evans says that no *notional* report can be given of what the subject in the steel-ball example is supposed to believe, but in a footnote he entertains (a) allowing that the subject does have a belief that admits of a relational specification, and (b) restricting "the application of [Russell's] Principle to thoughts which do admit of not 'He believes that the ball he is thinking of is pretty' a "notional" specification?

"make" the proposition true only in the sense that he would not know how to *verify* it; that is, he would have no capacity for determining the truth or falsity of the proposition. But, if this degree of verificationism is to show that the subject has no thoughts about the ball, then it ought to have received considerably more defense than Evans gives to it.

Russellian thoughts are typically information-based thoughts. A Russellian thought involving an idea of an object is informationbased when information or misinformation received from the object is within "the controlling conception" of thoughts involving that idea. And "a bit of information (with the content Fx) is in the controlling conception of a thought involving a subject's Idea of a particular object if and only if the subject's disposition to appreciate and evaluate thoughts involving this idea as being about an F thing is a causal consequence of the subject's acquisition and retention of this information" (122). Many interesting insights and suggestions are brought to bear in the initial discussion of information-based thoughts,5 but the main theoretical interest seems to be its bearing on an account of the structure of an idea of an object (here an appendix by McDowell is especially helpful). Such an account must discern (i) the information that controls thoughts that involve the idea, (ii) the object from which that information causally derives, and (iii) the mode of identification—the way of knowing which object the thought concerns—that the idea determines. An information-based thought is about a particular object just in case that object is both the object causally responsible for the relevant information and the object located by the mode of identification.

The foregoing apparatus provides the structure for a three-chapter, 158-page discussion of three kinds of modes of identification: demonstrative identification, self-identification, and recognition-based identification. In each case, Evans is concerned to show how ideas that determine these modes of identification enable thoughts involving them to satisfy Russell's principle. The account of recognition-based identification is somewhat vitiated by the doubt, expressed by McDowell in an appendix (but one the reader is apt to have before then), that recognition-based thinking about an object fails to constitute an autonomous mode of identification. The lively

⁵ I especially liked Evans's argument to show that there is "a logical gap between a thought that can be expressed in the words 'The Φ is F' and an information-based thought such as would be expressed in the words 'That Φ is F', even if the mode of identification employed in the latter exploits the fact that its object is uniquely Φ " (135).

chapter on self-identification will be of great interest to anyone exercised by 'I'-thoughts, but, because Evans is interested in this mode of identification mostly for its own sake, the chapter "represents something of a detour from the main line of argument" (137). The intricate and fascinating chapter on demonstrative identification, in which Evans tries "to make clear how demonstrative identification relates to the fundamental identification of material objects" (114), may be the best in the book. When a subject has a demonstrative thought about an object, he is informationally linked to the object by perception and able thereby to locate it in his egocentric space. Evans shows how the subject is able to have general knowledge of what makes propositions of the form $\lceil \pi = p \rceil$ true, where π is a fundamental idea of a position in public space and p an idea of a place in egocentric space. In this way a subject can be shown "to know what it is for This = the object at π now to be true (for arbitrary π)" (171), and, as the right-hand idea is fundamental, this explains how demonstrative Ideas satisfy Russell's principle.

I wish to call attention to the way in which thoughts involving a demonstrative idea, or mode of identification, of an object, such as a thought whose content is a proposition of the form This = the object at π now. This illustrates my earlier complaint about Evans's failure to provide a non-question-begging completion for the schema [*], but I want to use it now to raise a different problem for Evans's account of belief content in terms of modes of identification.

Consider again the thought that I think about Al when I think that he is drunk. According to Evans, the content of this thought involves Al, drunkenness, and a "mode of identification of," a "way of thinking about," Al, and it might, I suggested, be represented as the proposition $\langle M_{\rm al}$, drunkenness \rangle . Now $M_{\rm al}$, we know from the gloss in terms of [*], enjoys a specification that results from a replacement for 'R' in the form 'being R to Al', and the question arises what that specification might be. What, that is to say, is $M_{\rm al}$? It seems to me that, if the content of my thought really does involve a mode of identification, then it ought to be possible to say what that mode of identification is, and to say this without use of the technical term 'mode of identification' (for that technical term was introduced to mean merely whatever completes the content of a Russellian thought once one has been given the object thought about and what is thought about it). Yet Evans has provided no way of specifying this mode of identification which does not make essential reference to other thoughts involving it. This does not mean that Evans is giving a circular definition of 'mode of identification'; for he need not be

viewed as offering any definition of the notion at all. He is best understood, I think, as giving a partial explanation of modes of identification in terms of the functional relations that thoughts involving them stand in to other thoughts involving them. In that case, we would have the right to wonder whether there really are "modes of identification"—properties which are individuative of belief content and which have the functional roles Evans hopes to ascribe to them. And, if it should prove impossible to say in non-question-begging terms what these modes of identification are, then we should conclude that there are none. Evans, I believe, has given us no reason to think that there are content-involving properties that have the functional roles he has characterized.

LANGUAGE

Most uses of singular terms are Russellian (see above). Among the exceptions are "descriptive names," such as a name introduced by some such stipulation as "Let us call whoever invented the zip 'Julius'" (31). Definite descriptions are not exceptions, because in their non-Russellian use they are best treated not as singular terms but as quantifiers. The thesis that most uses of singular terms are Russellian is, of course, practically a corollary of the thesis that singular thoughts are Russellian: information-based thoughts are Russellian, and most uses of singular terms occur in utterances that cannot be understood without the audiences' having information-based thoughts about the referents of those singular terms.

Although most singular terms are Russellian, Evans does not subscribe to any 'Fido'-Fido semantics for them. It is quite possible for 't is F' and 't' is F' to say different things when 't' and 't' are Russellian singular terms having the same referent: utterances of the two sentences may require different kinds of information-based thoughts about the referent. There is, however, an important difference in this regard between proper names and other kinds of singular terms: to understand an utterance of 'NN is F', it is not required that an audience think of the referent of the name 'NN' in any particular way, and 'the single main requirement for understanding a use of a proper name is that one think of the referent' (400). This gives the appearance that Evans must hold that 'Phosphorus is F' and 'Hesperus is F' have the same content, but this does not cohere with his sympathetic discussion of Frege, and McDowell notes that Evans had planned to discuss the question.

An account according to which singular terms are Russellian should explain how something true can be said in the utterance of a negative existential statement of the form 't does not exist', when 't'

in its other uses is Russellian. There is a long chapter on this in which utterances such as 'That little green man does not exist' get analyzed in terms of a certain kind of make believe, or mutual pretense, engaged in by the speaker and hearer. Evans's account is quite ingenious, but seems not to have any application to negative existentials of the sort that ought to be most worrying to a Russellian. I have in mind the truth that would be conveyed by a sentence such as 'Qaddafi does not exist and never has existed: there is no such person' in a possible world just like ours but for the truth of that utterance. Evans's account fails to apply to such an utterance, because the truth conditions of the imagined utterance would surely make no reference to any pretense or make believe engaged in by the speaker and his audience.

The Varieties of Reference contains an introduction and has the following divisions: Part One, Historical Preliminaries, has the chapters "Frege," "Russell," and "Recent Work" (viz., Kripke); Part Two, Thought, has the chapters "Russell's Principle," "Information, Belief, and Thought," "Demonstrative Identification," "Self-identification," and "Recognition-based Identification"; and Part Three, Language, has chapters on "Communication and Information," "Existential Statements," and "Proper Names." Several chapters have appendixes by McDowell. This is a difficult book, but, for those concerned with its topics, well worth the effort to master it.

STEPHEN SCHIFFER

University of Arizona

On the Plurality of Worlds. DAVID LEWIS. Oxford: Basil Blackwell, 1986. ix, 276 p. Cloth \$39.95, paper \$12.95.*

David Lewis is well known for his brand of modal realism. In particular, he is well known for the extremeness of his brand of modal realism, and has earned himself some popular fame as an iconoclast: although he modestly calls his view just "modal realism" (2), it would

0022-362X/88/8501/0042\$00.50

© 1988 The Journal of Philosophy, Inc.

^{*} I thank Tak Yagisawa for his 1986 seminar at the University of North Carolina, and Alvin Plantinga for a stimulating symposium at the 1986 Eastern Division meeting of the American Philosophical Association, both of which helped me to formulate the ideas in this review.

be better termed "Mad-dog modal realism," or "Rape-and-loot modal realism," or "Nuclear-holocaust modal realism." In the present book, based on his 1984 John Locke Lectures, Lewis undertakes to defend his extreme realism against less daring forms. Unlike some contemporary philosophers who have made reputation and money by putting forward outrageous positions unbacked by even halfway convincing defense, Lewis scrupulously compares his view to alternatives and tries to show by the most honest and rigorous argument that it is preferable on the whole.

Lewis is, of course, a possible-worlds semanticist. But his shocking main thesis is that "other possible worlds" are just that, other flesh-and-blood worlds; they are not this-worldly mockups or substitutes such as stories, big sets, mental constructs, or other abstractions doing duty for "worlds" in modal theory. Nonactual possibilia exist, in just the same sense as we and our friends and our familiar possessions exist. This is why Lewis qualifies as an extremist; for one can perfectly well believe in the firm reality of modal properties and modal facts without acquiescing in the belief that nondenumerable hordes of nonactual but grittily physical entities and worlds exist just as we do.

So far we might be tempted to write off Lewis's extremism as terminological, giving him the word 'exist'. But there is more. According to Lewis, nonactuals differ from actuals only indexically, in not being worldmates of *ours*. To be "nonactual" is just to inhabit a world distinct from this one, though that alien world physically blooms and buzzes just as ours does and is in no ontological way subordinate. Thus, great importance attaches to the boundaries between worlds.

How then are worlds to be individuated? Lewis answers that what separates our world from the others is simply spatiotemporal disconnection: "things are worldmates iff they are spatiotemporally related. A world is unified, then, by the spatiotemporal interrelation of its parts" (71). Lewis toughs out (72/3) the obvious objection that it seems possible for a single world to contain a pair of spatiotemporally disconnected proper parts—by denying that possibility. But, granting the potential weirdness of actual and possible spacetimes, he concedes (74–78) that worldmatehood might be determined by only "analogically spatiotemporal" external relations.

The flesh-and-blood concreteness of Lewis's worlds comes, I think, from taking seriously Meinong's doctrine of *Sosein*: the golden mountain is golden and is a *mountain*, not a set or a proposition or a mental entity. And the flesh-and-bloodiness of the worlds

also accounts for some of Lewis's other views which might seem odd on their face: if worlds are physical universes, like planets only bigger and more drastically separate from each other, then it is fairly plain why Lewis thinks that (i) "actuality" is indexical in the way described;1 (ii) possibilia are grouped naturally into worlds in the way described and the spatiotemporal criterion is invoked; (iii) objects can only be counterparts, not genuinely identical, "across" worlds (ch. 4); (iv) there are no impossible objects or worlds (151);² and (v) worlds are disanalogous to times, particularly as regards persistence of individuals through them (202-204).

As I have said, Lewis's thesis is that of a modal plenum: "unrealized" possibilities are not really unrealized, but are only spatiotemporally or analogically-spatiotemporally cut off from us. Consider one of Lewis's own examples: although there are not, never have been, and never will be any talking donkeys, it is logically and metaphysically possible for there to have been talking donkeys. In virtue of what does this undoubted possibility obtain? Lewis claims that it obtains in virtue of there existing, in the absolute ordinary, everyday, and physical sense of 'exist', at least one donkey that has two extraordinary physical properties: (a) it talks; and (b) it is spatiotemporally or analogically-spatiotemporally disjoint from us.

Now, I think that both (a) and (b) are grotesque, or at least that no one will ever have the slightest evidence for believing either of them.3 If the foregoing target sentence is more or less obviously true, and its proposed analysans is indeed garishly false, then Lewis's theory is seriously embarrassed.

A related point is made by Brian Skyrms: 4 "If possible worlds . . . are supposed to exist in as concrete and robust a sense as our own . . . , then they require the same sort of evidence for their existence as [do] other constituents of physical reality" (326); there is not the slightest physical evidence for the existence anywhere of donkeys that talk. Lewis replies to Skyrms that "[o]ur knowledge can be divided into two quite different parts" (111): contingent knowledge

¹ But see also pp. 93/4.

² But see note 3 to p. 7, in which Lewis defends his rejection of impossibilia differently. I do not accept that defense, but I shall have to let it go for now. As Lewis notes and deplores (98/9), I used to complain that his view was only dubiously intelligible [e.g., in "The Trouble with Possible Worlds," in M. Loux, ed., The Possible and the Actual (Ithaca, NY: Cornell, 1979), p. 290)]. I hereby withdraw that complaint utterly; what Lewis is saying in the present work is now all too

^{4 &}quot;Possible Worlds, Physics and Metaphysics," Philosophical Studies, XXX (1976).

dd

er

iin

le-

ay

cts

ds

(v)

ce

al-

n-

er

ve

a-

ıe

it

y-

10

n-

10

ts

is

n

e

e

1S

e

requires causal contact with its objects, but "[o]ur necessary knowledge . . . requires no observation of our surroundings, because it is no part of our knowledge of which possible world is ours and which possible individuals are we" (112). For the case of mathematical and modal knowledge, we make "imaginative experiments," and "thereby cover an infinite class of worlds all in one act of imagining" (114). Lewis invokes a previously defended (though qualified) "principle of recombination," to the effect that "anything can coexist with anything else, [and] . . . anything can fail to coexist with anything else (88); the idea seems to be that we imagine recombinations of actual things, and infer that such recombinations are possible, hence(?) exist at an alternative world. I suspect that as a piece of psychology this conjecture is right. But I do not see that it, or the mental process it describes, provides anyone with evidence of the physical existence of talking donkeys spatiotemporally disconnected from us.

I have suggested that Lewis's modal metaphysics is wildly false. But Lewis himself admits its drawbacks, and claims only that competing accounts of possibilia are equally unsatisfactory. He rejects any program that explicates nonactual possibilia and all the phenomena of modality exclusively in terms of actual objects and their properties. In particular, he attacks "the ersatzist programme."

Actualist programs offer a good deal. They ostensibly accommodate the phenomena of modality, folding those phenomena into our overall account of the universe, and they do so wholly in terms of actual states of affairs, without reliance on Meinongian possibilia taken as primitive. Lewis admits these virtues and even grants that "the ersatzers have the advantage in agreement with common sense about what there is" (140), but still he calls actualism an ill-advised grab at "paradise on the cheap" whose seekers in fact end up "pay[ing] a high price," and in his chapter 3 he tries to defeat it.⁵

He distinguishes three forms of ersatzism: "linguistic," "pictorial," and "magical"; but of these he respects only the first, which construes "possible worlds" as, or replaces them by, sets of (somehow) linguistic entities. Lewis offers two main objections to linguistic ersatzism. The first is that any set of sentences or propositions purporting to describe a possible world must be a consistent set, and consistency is a matter of possible truth; for this and several other

⁵ Lewis seems to identify actualism with ersatzism. I would have called Kit Fine a non-ersatzing actualist (see Fine and A. Prior, *Worlds, Times, and Selves* (London: Duckworth, 1977)), but let it pass for now.

reasons (151–157) the linguistic ersatzer is stuck with a primitive modal notion and cannot claim to have explicated modality in non-modal terms.

The question of primitive modality is tricky. I know of no actualist who has been able to dispense with it (my own modal primitive is 'compatible' as applied to pairs of properties). But, if every actualist is stuck with some modal primitive, so, I would say, is Lewis. The flesh-and-bloodiness of his worlds might be thought to relieve him of the need for the sort of abstraction indulged in by actualists, and so it does; but Lewis mobilizes a modal primitive nonetheless. It is 'world'. 'World' for him has to mean "possible world," since the very flesh-and-bloodiness aforementioned prevents him from admitting impossibilia. Some sets of sentences describe "worlds" and some (the inconsistent ones as we know them to be) do not; but Lewis cannot make that distinction in any definite way without dragging in some modal primitive or other. Thus, actualists are no worse off in that regard.

Lewis's second objection is that the linguistic ersatzer cannot distinguish possibilities that are, in fact, distinct possibilities. In particular (159–165), there might have been alien "natural properties," properties which do not actually exist but do adorn worlds having (e.g.) an alien physics. Since those properties do not exist here, we have no way to name them, and so we cannot describe worlds in which they do exist, worlds in which they are switched around, etc.

Lewis has in mind what he calls "sparse" properties (60), the genuine physical properties which make for real similarity and which make a physical world go. It is easy to suppose that other worlds' fundamental particles might have unusual spins or "flavors" or the like, unknown in the actual world. But let us recall that, for Lewis, properties, sparse or not, are just sets of this-wordly and/or otherwordly individuals; such sets exist and have their members independently of what world one finds oneself in. Thus, the "alien" properties are only here-uninstantiated properties, i.e., for us, uninstantiated properties. There is nothing peculiar about that, at least for anyone who countenances properties in the first place.

Lewis anticipates the latter point and responds to it (160/1) by asking "what does the ersatzer mean by 'property'?" An ersatzer's "properties" are usually taken as primitive, since they are needed in the building of ersatz worlds to begin with. One might observe that

⁶ Stewart Shapiro and I have briefly but inadequately tried to address it in "Actuality and Essence," *Midwest Studies in Philosophy*, XI (1986): 358.

the notion of a property is intuitively more familiar than that of a possible world or that of a physical but nonactual object, and so it is less offensively taken as primitive. But uninstantiated properties are especially problematic, in that they cannot be abstracted from intances but are only "of a kind with" the better-understood properties that have been abstracted from actual instances. Lewis goes on (189–191) to complain that such properties can be supposed to "do their stuff" only by magic. I am not yet sure how much more serious a charge this is than just that they are indeed taken as primitive.

Lewis's book is far richer than I have been able to indicate here. In chapter 2, he confronts many more tough objections and wrestles them to the ground. ("I am mounting a defensive operation, and will be content with a standoff" (112); in a large number of cases, Lewis is forced to grant what he admits are implausibilities.) Chapter 4 deals with trans-world identity, or rather trans-world counterparthood, and contains in particular an intricate attack on David Kaplan's and R. M. Adams' haecceitism. The book simply bursts with illuminating argument, and is non-negotiably "must" reading for any metaphysician.

WILLIAM G. LYCAN

University of North Carolina

tive

on-

llist

e is

llist

The

1 of

o it

ld'.

sh-

im-

the

not

me

hat

discu-

s,"

ing

in itc. ihe ich ds' he re-en-er-en-

by r's in at

tu-

^{7 &}quot;How to Russell a Frege-Church," this JOURNAL, LXXII, 19 (Nov. 6, 1975): 716-729.

⁸ "Primitive Thisness and Primitive Identity," this JOURNAL, LXXVI, 1 (1979): 5–26; also, Lycan and Shapiro, op. cit.

NOTES AND NEWS

The editors are pleased to announce the Summer School & Conference on Mathematical Logic organized by Sofia University, and dedicated to the 90th anniversary of Arend Heyting, to be held September 13–23, 1988, in Chaika, Bulgaria. The scientific program will include lectures, discussions, and informal seminars on the following topics: recursion theory, modal and nonclassical logics, intuitionism and constructivism, related applications to computer science, and the life and work of Heyting. More information may be obtained by writing to: Heyting '88, Sector of Logic, Mathematics Faculty, Boul. Antoa Ivanov 5, Sofia 1126, Bulgaria.

Effective Volume 20 (1987), APEIRON, a journal for ancient philosophy and science, has moved from Monash University, Australia to the University of Alberta, Canada. The new Executive Editor is Roger A. Shiner, and the other members of the Editorial Board are Richard Bosley, John King-Farlow, Mohan Matthen, Francis Jeffry Perretier, Janet Sison, and Martin Tweedale. The address for the new editorial office is: Department of Philosophy, University of Alberta, 4-108 Humanities Centre, Edmonton, Alberta, Canada, T6G 2E5. The new publisher is Academic Printing and Publishing, Box 4834, Edmonton, Alberta, Canada T6E 5G7, to whom should be sent all subscription and business correspondence.

The International Philosophers for the Prevention of Nuclear Omnicide are pleased to announce that they will present symposium sessions at the XVIIIth World Congress of Philosophy, August 20–27, 1988, in Brighton, England. The topic will be "Philosophy of Peace in the Nuclear Age: What Must We Do to Prevent Nuclear Omnicide?" Papers or abstracts may be sent to: Secretariat, IPPNO, 1426 Merritt Drive, El Cajon, CA 92020. IPPNO is glad to report that it has been chosen by the United Nations as one of the organizations to receive its special Peace Messenger Award "in recognition of a significant contribution to the programme and objectives of the international Year of Peace." This was a tribute to all who took part in IPPNO's First International Conference in St. Louis, which was jointly sponsored by the Concerned Philosophers for Peace and the Society for the Philosophical Study of Marxism, and which was held in conjunction with the 1987 annual meeting of the American Philosophical Association, Central Division.

The Department of Philosophy at the University of Bristol will again be able to host an overseas philosopher in 1988/89. We can offer an office, help in finding accommodations, library rights, and participation in department activities. Philosophers may contact Adam Morton, Department of Philosophy, 9 Woodland Road, Bristol BS8 1TB, England.

The London School of Economics and Political Science is pleased to announce that the second Lakatos Award has been awarded to Michael Friedman (University of Illinois, Chicago) for his book *Foundations of Space-Time*

Theories, and Philip Kitcher (University of California, San Diego) for his book Vaulting Ambition: Sociobiology and the Quest for Human Nature. The Award, which is for an outstanding contribution to the philosophy of science, has been endowed by the Latsis Foundation in memory of Imre Lakatos and will be awarded annually for at least ten years.

e on

0th

ika, for-

assi-

uter

ned

ntoa

and y of

her

OW,

ale.

Jni-

ıda,

34,

rip-

are

Ith

nd.

Do

tar-

to

izaf a

nal

irst

the

ical

ual

ble

in ac-

hy,

an-

ed-

me

The editors of the *Philosophical Inquiry*, Monograph series announce the publication of their first monograph, *The Concept of Causality in Presocratic Philosophy*, by D. Z. Andriopoulos. They also invite philosophers and others concerned to submit their research work for publication to *Philosophical Inquiry* Monographs, P.O. Box 84, Aristotelian University of Thessaloniki, Greece.

The University of St. Andrews Centre for Philosophy and Public Affairs will hold its third International Conference September 15–17, 1988. Subjects will include: the ontology of morality, method in ethics, rationality and decision theory, morality and motivation. Principal participants will include: John Broome, Jonathan Dancy, James Griffin, Shelley Kagan, Larry Temkin, and Susan Wolfe. Papers are invited for subsidiary sessions, and abstracts must be submitted by March 31, 1988 to: Gordon Graham, Centre for Philosophy and Public Affairs, Department of Moral Philosophy, University of St. Andrews, KY16 9AL, Fife, Scotland.

The editors are pleased to announce *Applied Philosophy*, a new book series to be published by Peter Lang Publishing under the editorship of Farhang Zabeeh. They invite original manuscripts in applied philosophy in which semantic analysis is used for clarification or solution of contemporary moral, political, social, and religious issues. Inquiries may be sent to: F. Zabeeh, Department of Philosophy, Roosevelt University, Chicago, IL 60605.

The International Berkeley Society is pleased to announce a group meeting to be held conjointly with the Eastern Division meeting of the American Philosophical Association in Washington, DC, December 27–30, 1988. The purpose of the meeting is to feature the work of new Berkeley scholars. For purposes of the meeting, "new Berkeley scholars" will be those persons who are not yet six years past the date of receipt of the PhD and persons who have not yet received the PhD. Papers and biographical information may be submitted (by July 1, 1988) to: A. David Kline, Philosophy Department, Iowa State University, Ames, IA 50011.

The Philosophy Department of the California State University/Fullerton will sponsor its 18th annual philosophy symposium on February 24–27, 1988. The topic will be: Moral Reasoning in China. This symposium will focus on conceptions of virtues, moral character, and moral decision making in Chinese philosophy and culture. Speakers will include: David Wong (Brandeis), Antonio Cua (Catholic University), Kwong Loi Shun (Berkeley), Chad Hansen (University of Vermont). Further information may be obtained by writing

to: Craig K. Ihara, Philosophy Department, California State University, Fullerton, CA 92634.

The Council for Philosophical Studies, with support from the National Endownent for the Humanities, will present the following Institutes this summer: (1) "Early Modern Philosophy," June 27-August 5, 1988 at Brown University, Providence, Rhode Island. The purpose of this institute is to expose teachers of philosophy to recent work in the history of early modern philosophy (roughly 1600 to 1750), and help them to integrate new material and approaches into their teaching and research. Institute staff will include: Daniel Garber (Director), Roger Ariew, Michael Ayers, Jonathan Bennett, Edwin Curley, Alan Gabbey, Jean-Luc Marion, David Fate Norton, Richard Popkin, Jerome Schneewind, Robert Sleigh, and Margaret Wilson. Further information and application forms may be obtained by writing to: Daniel Garber, NEH Summer Institute, Department of Philosophy, University of Chicago, 1050 East 59th Street, Chicago, IL 60637. (2) "Aristotle's Metaphysics, Biology, and Ethics," directed by John Cooper, Michael Frede, and Allan Gotthelf. This institute will be held at the University of New Hampshire, Durham, NH, June 20-July 29, 1988. More information may be obtained by writing to Allan Gotthelf, Department of Philosophy, Trenton State College, NJ 08625. And (3) "Interpretation in the Sciences and Humanities," to be held June 20-July 29, 1988 at the University of California/Santa Cruz. The institute staff will be: Stanley Cavell; Hubert Dreyfus, Co-Director; Clifford Geertz, Institute for Advanced Studies; David Hoy, Co-Director; Thomas S. Kuhn; Alexander Nehamas; Richard Rorty; and Charles Taylor. Problems to be debated include whether the practices of interpretation in various fields can be captured in a single comprehensive theory, whether interpretive practices compete with supposedly objective, cultural-neutral, quantitative methods, whether the reading of texts is an alternative paradigm in the social and human sciences, and whether some interpretations can be shown conclusively to be better than others. More information may be obtained by writing to: David Hoy, Kresge College, University of California, Santa Cruz, CA 95064. The Council is also pleased to announce the availability of a paperback book, Teaching Theory of Knowledge, based on the summer Institute on Theory of Knowledge which was directed by Alvin Goldman and Keith Lehrer. It contains information and advice about teaching theory of knowledge and offers several sample courses as well as reference material. The volume, which was edited by Marjorie Clay, is available to individuals without charge due to support for the project from the NEH. A copy may be obtained by writing to: Alan Mabe, Project Director, 203 Dodd Hall, Florida State University, Tallahassee, FL 32306.

The National Endowment for the Humanities will again be offering seminars for college teachers during the summer of 1988. The philosophy seminars will be: "Consequentialist Theories of Morality," "Justice and Health Care," "Issues in the Philosophy of Childhood," "Frege and the Philosophy of Mathematics," and "The Philosophy of Socrates." Applications to this year's

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

seminars (due March 1, 1988) and information about the program for 1989 may be obtained by writing to: Summer Seminars for College Teachers, Room 316, Division of Fellowships and Summer Seminars, National Endowment for the Humanities, 1100 Pennsylvania Avenue, Washington, DC 20506.

Ful-

En-

um-Own

to

ern

rial

ide:

ett,

ard

her

niel of

etaand

np-

ob-

ate

ni-

nta

or;

or;

or.

in

ner

al, gm

be

b-

ia,

oil-

er

nd

of

al.

als

be da

rs

rs

The Newberry Library invites applications for Resident Fellowships in the Humanities for 1988–89. Terms in residence may be as short as a few weeks or as long as eleven months; stipends fall in an equally broad range. Special awards, all for periods of six to eleven months, include National Endowment for the Humanities Fellowships, the Monticello College Foundation Fellowship for women, and (new this year) the Lloyd Lewis Fellowship for senior scholars in American History. Further information and application forms may be obtained by writing to: Awards Committee, Newberry Library, 60 W. Walton St., Chicago, IL 60610.

The Charles S. Peirce Sesquicentennial International Congress, supported by the National Endowment for the Humanities, will be held at Harvard University the week preceding 10 September 1989 (the one-hundred-fiftieth anniversary of Peirce's birth). The Congress Organizing Committee hereby issues a call for papers on any aspect of Peirce's work, in any of the several fields in which he conducted research (science, mathematics, philosophy, history of science, logic, phenomenology, cosmology, semeiotic). The theme of the congress is Peirce and Contemporary Thought. Three copies of a 200-word, double-spaced abstract (two copies without name) may be sent to: K. L. Ketner, Chairperson, Organizing Committee, Institute for Studies in Pragmaticism, 304K Library, Texas Tech University, Lubbock, TX 79409.

Social Theory and Practice plans to publish a special issue on "Marxism and Feminism: Powers of Theory/Theories of Power," under the guest editorship of Roger S. Gottlieb and Nancy Holmstrom. We welcome critical (not historical or exegetical) papers on all aspects of the present and future of Marxist and feminist theory. Among the topics that might be considered are the compatibility or incompatibility of the two theories, problems they share, visions of a socialist/feminist society, and critical analyses of Marxist/feminist thinkers. Papers should be submitted by June 1, 1988 to Social Theory and Practice, Department of Philosophy, Florida State University, Tallahassee, FL 32306.

The University of Iowa is pleased to announce a conference on the topic "Ideas: Sensory Experience, Thought, Knowledge and their Objects in 17th and 18th century Philosophy." The conference will be held April 6–9, 1989. Sponsoring and endorsing organizations include the International Berkeley Society, the Hume Society, the Leibniz Society of America, the North American Kant Society, and Iowa State University. Scholars are invited to submit

papers on topics relating to one or more of the conference themes. Submissions should be sent by July 1, 1988 to Phillip Cummins, Department of Philosophy, University of Iowa, Iowa City, IA 52242.

The Department of Philosophy of the University of North Carolina/Greensboro is pleased to announce its twelfth Annual Symposium in Philosophy, to be held April 15–17, 1988. The topic is "The Future of Folk Psychology," and speakers include Paul Churchland (UC/San Diego), Stephen Stich (UC/San Diego), Daniel Dennett (Tufts), Jonathan Bennett (Syracuse), and Simon Blackburn (Oxford). Commentators include John Greenwood (UNC/Greensboro), John Heil (Davidson), Jeff Coulter (Boston), and Jay Rosenberg (UNC/Chapel Hill). Further information and registration forms may be obtained from the Department of Philosophy, University of North Carolina, Greensboro, NC 27412–5001.

The Department of Philosophy at the University of Oklahoma is pleased to announce the establishment of the biennial David Ross Boyd Lectures, named in honor of the first president of the University. The inaugural lectures will be delivered by Donald Davidson under the general title "The Objectivity of Values." The lectures will take place in Norman on February 1, 3, and 5, 1988. The Department is also pleased to announce that Neera Badhwar and Charles Mills have been appointed as assistant professors beginning September 1987.

The Department of Philosophy of Clark University is pleased to announce the appointment of Judith Wagner DeCew as Assistant Professor effective September 1987.

The Department of Philosophy at Florida State University is pleased to announce that Michael D. Bayles has joined the department with the rank of Professor and Charles David McCarty has joined the department as an Assistant Professor with a coordinated appointment in Computer Science. Also, Jaakko Hintikka has been appointed to the university's Foundation Professorship in Philosophy.

The Department of Philosophy of Georgia State University is pleased to announce the appointment of David Blumenfeld as Professor of Philosophy and Department Chair.

The Department of Philosophy of Syracuse University is pleased to announce two recent appointments to its faculty: Michael Stocker joins us as Irwin and Marjorie Guttag Professor of Ethics and Political Philosophy beginning in January 1988. Brent Mundy joined us as Assistant Professor of Philosophy in September 1987.

The Department of Philosophy of the University of Iowa is pleased to announce the appointments of Guenter Zoeller as Assistant Professor, effective August 1987, and David Stern as Assistant Professor, effective August 1988.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

The editors cannot afford to send to the printer manuscripts that are difficult for the compositor to read or footnotes that need to be completed in proof. They ask that manuscripts submitted be prepared according to the following guidelines:

- 1. All material, including the quotations and footnotes—especially footnotes—must be typed, double-spaced, on one side of the paper, with ample margins left and right, top and bottom.
- 2. No bibliographies will be published; bibliographical material should be put into the footnotes.
- 3. Footnotes should be complete, according to *Journal* style (see any issue). Journal citations should include name of journal in full, volume number, issue number, month or season, year, and pages. Book references should include publisher, city of publication, and date. Footnotes should be gathered at the end of the article.
- 4. Quotations should be checked by the author; capitalization, italics, spelling, etc., should match the original. Corrections should not have to be made (expensively) in proof.
- 5. If there are diagrams, they should be provided camera-ready by the author. If there are special symbols, they should be widely spaced and clearly marked; if there are many, a list should be provided. Standard logic symbols are available, and letters of the greek and script alphabets; no boldface.
- 6. Computer copy can be extremely difficult to read; only letter-quality copy is acceptable. It should be double-spaced and not justified. The print should be dark enough and the letters widely enough spaced so that both letters and punctuation marks can be read easily.
- 7. Finally, photo copies on shiny paper are not acceptable. Ribbon copy is much preferred, but clear copies on bond paper are acceptable if necessary.

The editors are willing to read and evaluate manuscripts that do not meet these standards, provided they are genuinely legible and not unreasonably long (over 15000 words). They ask for revision of excessively long manuscripts, and they require retyping of accepted manuscripts, if necessary, before scheduling for publication.

The Journal of Philosophy is published monthly, in issues that (ordinarily) contain only 56 pages of copy; this limits the length of papers the editors can normally consider, to no more than 7,500 words, or eighteen JOURNAL pages. Manuscripts very much in excess of these limits will be returned unread.

The Journal of Philosophy does not publish papers that are primarily historical or expository; the editors urge that such manuscripts be submitted elsewhere Gurukul Kangri Collection, Haridwar

The editors and trustees of The Journal of Philosophy wish to draw to the attention of our readers the notice of permission for personal or educational copying which appears at the bottom of the masthead page of each issue of this JOURNAL.

In the last few years we have received many requests for permission to photocopy or mimeograph articles for classroom use; we have also received many requests from libraries for blanket permission to photocopy. And we are aware of course that a great deal of such photocopying has been done without permission. We are grateful to those of you who have asked for our permission.

The copyright law of 1976 grants certain new rights to libraries and educators who need to photocopy and certain new protections to copyright owners. In general, photocopies in excess of "fair use" copying cannot be made without authorization by the copyright owner. What constitutes "fair use" is not yet entirely clear, but some single-copying for personal use is permitted, and multiple-copying for classroom use is clearly limited: the article copied must be short (less than 2,500 words) or only an excerpt copied; the copying must be spontaneous (at the instance of the individual instructor, and needed so promptly as not to allow time to obtain permission); and each copy must include a notice of copyright.

The Association of American Publishers has been instrumental in setting up an organization, the Copyright Clearance Center, designed to make it as simple as possible for organizations to pay the fees required for copying beyond fair use. The Journal has joined the Center and thereby authorizes such copying under the terms established by the Center. The per-copy price for personal or educational copying beyond fair use appears on the first page of each article in the Journal. That price is 10¢/page, with a minimum of 50¢, unless some special arrangement is requested by the author. The Journal continues to give permission for multiple-copying for classroom use, but at 5¢/copy/page, where an instructor applies directly to us for our permission. We turn over to our authors all substantial sums (\$15 or more) obtained from their work.

Journal issues are available from this office, singly as well as in volumes, back to 1904; prepaid orders are filled by return mail.

JOHNSONIAN PRIZE TO PHILOSOPHY

The editors very much regret that the Johnsonian Prize in Philosophy will not be awarded in 1987. Previous Johnsonian prize winners have been: 1980, Daniel Bonevac, Reduction in the Abstract Sciences (Indianapolis: Hackett, 1981); 1981, Igal Kvart, A Theory of Counterfactuals (1986); 1982, Moreland Perkins, Sensing the World (1983); 1984, Robert J. Dunn, The Possibility of Weakness of Will (1986); and 1985, Harry van der Linden, Kantian Ethics and Socialism (forthcoming in 1988); 1986, Clyde L. Hardin, Color for Philosophers (1988). Also, honorable mention in 1981 went to Charles B. Guignon, Heidegger and the Problem of Knowledge (1983); and in 1986 to Dionysios Anapolitanos, Leibniz: Representation, Continuity, and the Spatiotemporal.

The Johnsonian Prize, which consists of one thousand dollars and guaranteed publication of the prize-winning book, will be offered again in 1988. The prize is designed to encourage the work of unpublished philosophers; authors are eligible who have published no book-length work. Books are selected from among manuscripts sponsored by departments of philosophy. Authors and departments who wish to submit manuscripts should observe the following guidelines:

1. Any graduate school of philosophy that confers the Ph.D. degree may submit one manuscript per year. It is not required that the work sponsored by a department be written by either a graduate or a faculty member of that department; it may, but need not be a revised version of a departmental dissertation. The closing date for this year's prize is August 1, 1988.

2. To qualify, the manuscript must be: (a) of book length (at least 40,000 words) and in proper form for publication as a book; (b) accompanied by an abstract of up to 2000 words; (c) written by someone who has not yet published any book-length work; (d) accompanied by two professional evaluations and a letter from the sponsoring department (the editors are depending on the departments to screen out all but the most important and original work); (e) uncommitted to a publisher. Manuscripts will be returned to the authors only if accompanied by appropriate postage.

3. Graduate schools who wish to participate should notify the JOURNAL in advance of submitting a manuscript for consideration, and, if possible, send

evaluations and abstracts in advance.

4. Preference will be given to the work of younger philosophers, especially those who have not yet been granted tenure.

Announcement of the seventh award will appear in the January 1989 issue of this JOURNAL; some manuscripts may be given honorable mention. Publication of the prize-winning book will be guaranteed by Hackett Publishing Company after the author has made what revisions are deemed necessary by the editors, if possible within twelve months of the award. The sponsors hope that the prize-winning books for both 1985 and 1986 will be on display at the APA Eastern Division meetings in 1988.

Digitized by Arya Samaj Foundation Chennai and eGangotri Chicag

THE EMBERS AND THE STARS

A Philosophical Inquiry into the Moral Sense of Nature

Erazim Kohák

"A richly evocative book. Kohák is a mature scholar, drawing from traditionally recognized philosophers, the Czechs Jan Patočka and Emanuel Radl, the American personalists Bowne and Brightman, and the impetus of Robert Frost."—Timothy F. Murphy, International Philosophical Quarterly

Paper \$11.95 284 pages

BAD ACTS AND GUILTY MINDS

Conundrums of the Criminal Law

Leo Katz

"Katz uses the insights of contemporary philosophical and psychological theory to enrich our understanding of how our substantative criminal law is and ought to be structured . . . Katz's book should become a classic in the field for years to come."-Michael S. Moore, author of Law and Psychiatry

Paper \$14.95 360 pages Library cloth edition \$45.00 Studies in Crime and Justice series

MEDIEVAL COSMOLOGY

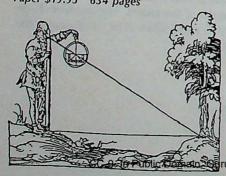
Theories of Infinity, Place, Time, Void, and the Plurality of Worlds

Pierre Duhem

Edited and translated by Roger Ariew

"Those who think that nothing happened in cosmology between Ptolemy and Copernicus badly need a course of reading in Duhem. Roger Ariew provides the opportunity in his admirable Medieval Cosmology."

-Edward Harrison, Sky and Telescope Paper \$19.95 634 pages



HUSSERL AND THE SEARCH FOR CERTITUDE

Leszek Kolakowski

With a new Preface

"[Kolakowski] places Husserl in the tradition of philosophers, from Descartes to the logical positivists, who were engaged in the attempt to discover some knowledge which was certain and indubitable." -Mary Warnock, Times Literary Supplement Paper \$8.95 96 pages

MARTIN HEIDEGGER

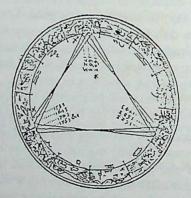
George Steiner

"It would be hard to imagine a better introduction to the work of philosopher Martin Heidegger." -George Kateb, The New Republic. Paper \$10.95 184 pages

THE EMERGENCE OF MORALITY IN YOUNG CHILDREN

Edited by Jerome Kagan and Sharon Lamb How-and when-do children distinguish right from wrong? These essays by prominent psychologists and a moral philosopher examine this issue.

Cloth \$24.95 360 pages 8 line drawings, 21 tables



KEPLER'S GEOMETRICAL COSMOLOGY

J. V. Field

Field examines Johannes Kepler's search for the geometrical plan according to which God created the universe. Field's sound scholarship provides a more complete picture of this enigmatic man and his work.

Cloth 32 offection of paragress 88 figures, 21 tables

THE JOURNAL OF PHILOSOPHY

LOGICAL AND ANALYTIC TRUTHS THAT ARE NOT NECESSARY*

THE notions of logical truth, analytic truth, and necessary truth are extremely important in philosophy. A logically true sentence remains true no matter what the interpretation of the nonlogical constants. An analytically true sentence is true in virtue of the meanings of its words. Necessarily true sentences are true in all possible worlds. The concepts involved in these definitions are central to philosophy, and it is of utmost importance that philosophers chart their interactions and examine where and how the distinctions among them evolve. Many philosophers agree that logical truths are analytic, since they are true in virtue of the fixed meanings of the logical constants. Examples like "All bachelors are unmarried" demonstrate that not all analytic truths are logically true, for though this sentence is true in virtue of the standard interpretation (meaning) of 'bachelor' and 'unmarried', there may be interpretations of these terms in which this sentence is false. Many philosophers also agree that there are necessary truths which are not logically true. Paradigm cases of such are identity statements of the form 'a = b' where 'a' and 'b' are both names (or logical constants). But many philosophers regard all logical and analytic truths as paradigm cases of necessary truths. After all, if a sentence is true solely in virtue of its logical form or in virtue of the meanings of its words, it seems that it could not fail to be true.

Saul Kripke just stipulates that analytic truths are necessary, in his lectures of 1970:¹

"Naming and Necessity," in Donald Davidson and Gilbert Harman, eds. (Boston: Reidel, 1972), pp. 252–355; reprinted independently (Cambridge, MA: Harvard, 1989)

vard, 1980).

^{*} This paper was written during my fellowship at the Center for the Study of Language and Information, which is funded by grants from the Systems Development Foundation. I would like to thank Jon Barwise, Ned Block, John Etchemendy, David Israel, Michael Jubien, Chris Menzel, Paul Oppenheimer, John Perry, and Chris Swoyer, for the discussions we have had on the topics contained herein.

Another term used in philosophy is 'analytic'. Here it won't be too important to get any clearer about this in this talk. The common examples, nowadays, are like 'bachelors are unmarried'. . . . At any rate, let's just make it a matter of stipulation that an analytic statement is in some sense true by virtue of its meaning and true in all possible worlds by virtue of its meaning. Then something which is analytically true will be both necessary and a priori (264).

Of course, Kripke is entitled to stipulate whatever he wants, but in this paper we plan to show that the following are consequences of the traditional conception of logical truth, analytic truth, and necessity: (1) there are sentences which are logically true but which are not necessary, and (2) there are sentences which are analytically true, true in virtue of the meanings of their words, yet which are not true in all possible worlds.

To argue for these consequences in a philosophically neutral way, we shall suppose that interpreted sentences, rather than propositions, are the bearers of truth. In section I, we establish that there are surprisingly simple examples of both logical truths that are not necessary (LTNNs) and analytic truths that are not necessary (ATNNs). In section II, we consider whether there are any ways to undermine the claims in section I. In this section, a certain definition of logical truth for modal languages, which has become established in the literature as an alternative to Kripke's original definition, is discovered to be incorrect. In section III, we investigate whether there are simpler examples of LTNNs and ATNNs than those described in section I. Finally, section IV contains a brief discussion of some important consequences that should be drawn.

T

We can find examples of LTNNs and ATNNs first-order modal language in which rigidly designating definite descriptions of the form $(1x)\phi$ are generated as primitive singular terms.² The following, si-

² Such a language is of considerable interest to a philosophical logician. It is the principal tool by which one can investigate the claim that certain definite descriptions of English can be adequately represented as rigid designators. Elsewhere, I have shown that a simple modification of this language can handle nicely a recalcitrant group of English descriptions; see Abstract Objects: An Introduction to Axiomatic Metaphysics (Boston: Reidel, 1983), pp. 99–106. This language is also interesting because it is immune to W. V. Quine's attempt to prove that modal distinctions collapse in ordinary quantified modal logic with complex singular terms Congress of Modal Involvement," Proceedings of the XIth International Paradox and Other Essays (New York: Random House, 1966), pp. 156–173]. Dagfinn Føllesdal has shown that Quine's argument does not apply to modal languages in which all (complex) singular terms are rigid ["Referential Opacity and Modal Logic," PhD thesis: Harvard University, 1961, p. 116; "Situation Semantics and the 'Slingshot' Argument," Erkenntnis, XIX, 1 (May 1983): 91–98].

multaneous recursive definition of term and formula defines such a language:

- 1. Every constant a_i and variable x_i is a term.
- 2. If R^n is any *n*-place predicate, and τ_1, \ldots, τ_n are any terms, then $R^n \tau_1, \ldots, \tau_n$ is a formula.
- 3. If ϕ and ψ are any formulas, and x_i is any variable, then $(\sim \phi)$, $(\phi \rightarrow \psi)$, $(\forall x_i)\phi$, and $(\Box \phi)$ are formulas.
- 4. If ϕ is any formula, and x_i is any variable, then $(ix_i)\phi$ is a term.

Since descriptions are generated as primitive, complex terms, they do not have scope. The notion of scope no more applies to them than it does to constants or variables. Since descriptions are not contextually defined, they cannot be eliminated, and so the question of eliminating them with wide or narrow scope does not arise. Intuitively, the symbol '1x' expresses the primitive logical notion "the x such that," just as ' \sim ' expresses the primitive logical notion "it is not the case that," and ' $\forall x$ ' expresses the primitive logical notion "every x is such that."

If the primitive descriptions are to designate rigidly, we have to interpret the language so that they denote that object at the distinguished actual world which uniquely satisfies the description (should there be one), even when the description occurs in the scope of a modal operator. This could be done by using the following semantic characterizations for the language. First, we specify an interpretation of the language to be something like that in Kripke's work.3 An interpretation $\mathfrak A$ is a set $\langle W, W_0, D, F \rangle$, where W is a nonempty set (of possible worlds), Wo is a distinguished element of W (Wo is the base world), D is the domain of (possible) objects, and F is the function that assigns to each constant a_i of the language an element of D, and assigns to each predicate letter R_i^n a function from worlds to sets of n-tuples.4 An assignment f to the variables with respect to an interpretation I is then defined in the usual way. For convenience, we always drop the subscript that relativizes f to an interpretation, though all fs are to be understood as so relativized. In terms of these two notions, we may define the notions of denotation and satisfaction, which recursively call each other.

3 "Semantical Considerations on Modal Logic," reprinted in Leonard Linsky, ed., Reference and Modality (New York: Oxford, 1971), pp. 63-72.

It will not be crucial to my argument that the domain of individuals be nonempty. My examples of LTNNs do not depend on the nonlogical claim that there must be at least one individual in the domain of discourse. Also, for the purposes of this paper, predicates are assigned functions from worlds to sets of *n*-tuples, in the usual way. I prefer the formulation in my *Abstract Objects*, however, in which predicates denote primitive relations, the *extensions* of which are functions from worlds to sets of *n*-tuples. The definition of denotation is given first, and, in it, we use 0 as a metavariable ranging over the elements of D, and we use " $f \stackrel{*}{=} f$ " to mean that f' is just like f except perhaps in what it assigns to x. The denotation of term τ with respect to interpretation $\mathfrak A$ and assignment f (" $\mathfrak A_{\mathfrak A,f}(\tau)$ ") is defined as follows:

- 1. where τ is any constant κ , $d_{\mathfrak{A},f}(\tau) = F(\kappa)$
- 2. where τ is any variable ν , $d_{\mathfrak{A},f}(\tau) = f(\nu)$
- 3. where τ is any description $(\imath x)\psi$, $d_{\mathfrak{A},f}(\tau) =$

$$\begin{cases} o \text{ iff } (\exists f')(f' \stackrel{x}{=} f \& f'(x) = o \& f' \text{ satisfies } \psi \text{ at } W_0 \\ \& (\forall f'')(f'' \stackrel{x}{=} f' \& f'' \text{ satisfies } \psi \text{ at } W_0 \rightarrow f'' = f')) \\ \text{undefined, otherwise} \end{cases}$$

Clause (3) says that for $(1x)\psi$ to denote an object 0, there must be a certain assignment to the variables which uniquely satisfies ψ at the actual world. Intuitively, this guarantees that the denotation of a description $(1x)\psi$ is just the unique object 0 that the description picks out at the actual world, should there be one. This packs Russell's analysis of the description into the semantic conditions that must obtain for the description to denote an object. Note that this definition of denotation defines a unary function on the terms of the language. All the terms receive denotations simpliciter; they do not receive denotations relative to each world, for this is unnecessary when all terms rigidly designate.

We need next the notion of satisfaction, and it will suffice to state only the base clause and modal clause of this definition, since the others are routine. There is one little twist to the definition of satisfaction, which prevents the base clause from being undefined for formulas that may have nondenoting descriptions. Typically, the base clause for atomic formulas ϕ of the form $\rho^n \tau_1 \ldots \tau_n$ reads as follows: f satisfies ϕ at world w iff $\langle d_{\mathfrak{A},f}(\tau_1), \ldots, d_{\mathfrak{A},f}(\tau_n) \rangle \in [F(\rho^n)](w)$. But this clause would be undefined for formulas containing descriptions that failed to denote. So the clause requires a minor modification, which is accomplished by the following, general definition:

If given an interpretation $\mathfrak A$ and assignment f, we define f satisfies ϕ at world w as follows:

1. Where ϕ is an atomic formula of the form $\rho^n \tau_1 \dots \tau_n$, f satisfies ϕ at W iff $(\exists o_1) \dots (\exists o_n)(o_1 = d_{\mathfrak{A},f}(\tau_1) \& \dots \& o_n = d_{\mathfrak{A},f}(\tau_n) \& \langle o_1, \dots, o_n \rangle \in [F(\rho^n)](W))$

2. . . .

This base clause for satisfaction simply stipulates that, if an assignment f is to satisfy an atomic formula ϕ at a world w, every term in ϕ must have a denotation (simpliciter) and the denotations must exhibit the appropriate structure at w. One should check that the definition works as it should by considering a few basic cases. Note that the definition of satisfaction and the definition of denotation recursively call each other, and that the interaction yields the correct satisfaction conditions for formulas containing descriptions. These two definitions are not circular, despite appearances. By fixing an arbitrary interpretation, and working through a simple example like "f satisfies $P(\imath x)Qx$ at W_0 ," one is able to derive satisfaction conditions stated solely in terms of set-theoretic membership (see the Appendix for the details). In other words, both notions are eliminable, and, consequently, there is no circularity.

The modal clause in the definition of satisfaction is perfectly standard: f satisfies $\Box \phi$ with respect to world w iff $(\forall w')$ (f satisfies ϕ at w'). This S5 interpretation without an accessibility relation keeps everything simple.

We are now in a position to state the definitions of truth and logical truth, and it will be seen that they are the ordinary, straightforward concepts with which we are all familiar. First, we say that ϕ is true under interpretation $\mathfrak A$ at world $\mathfrak A$ iff every assignment $\mathfrak A$ satisfies ϕ at $\mathfrak A$. Next we say, ϕ is true under $\mathfrak A$ iff ϕ is true under $\mathfrak A$ at $\mathfrak A$ 0. Finally, ϕ is logically true iff for every interpretation $\mathfrak A$, ϕ is true under $\mathfrak A$ 1 (compare Kripke, "Semantical Considerations," op.cit., p. 64).

These definitions reveal that there are both LTNNs and ATNNs. We can demonstrate this by considering a particular sentence of our language and establishing several important claims about it. The sentence that will play a central role in the next few pages is ϕ_1 :

$$(\phi_1) P(\imath x)Qx \rightarrow (\exists y)Qy$$

This is to be read: If the Q-thing is P, then something is Q. It should be relatively easy to see that the following four claims are true (some are straightforward consequences of the foregoing definitions).

Claim 1: ϕ_1 is logically true.

To see this, pick an arbitrary interpretation \mathfrak{A}_j and assignment f_k . Clearly, either f_k satisfies the antecedent of ϕ_1 at w_0 , or it does not. If f_k satisfies $P(\iota x)Qx$ at w_0 , then, by the definition of denotation, the description uniquely picks out an object at w_0 . If so, then f_k satisfies $(\exists y)Qy$ at w_0 (by the recursive clause for quantifiers in the definition of satisfaction). So, by the clause in the definition of satisfaction

governing conditionals, f_k satisfies ϕ_1 at w_0 . If f_k fails to satisfy the antecedent of ϕ_1 at w_0 , then f_k still satisfies ϕ_1 at w_0 . So ϕ_1 is logically true.

Claim 2: $\Box \phi_1$ is not logically true.

To prove this, we have to show that there is an interpretation under which $\Box \phi_1$ is not true. Consider the interpretation \mathfrak{U}_1 having the following characteristics: (1) there is an object that not only uniquely exemplifies Q at w_0 but also exemplifies P in every possible world, and (2) there is a world w_1 at which nothing exemplifies Q. Under \mathfrak{U}_1 , $\Box \phi_1$ is not true. The reason is that there is a world, namely w_1 , at which ϕ_1 is not true. ϕ_1 fails to be true at w_1 because the antecedent P(x)Qx is true there (by characteristic (1) of \mathfrak{U}_1), but the consequent \mathcal{U}_1 is false there (by characteristic (2) of \mathfrak{U}_1). Since there is an interpretation under which $\Box \phi_1$ is not true, $\Box \phi_1$ is not logically true.

Claim 3: Under \mathfrak{A}_1 , ϕ_1 is a logical truth which is not necessary.

By Claim 1, ϕ_1 is a logical truth, and so remains a logical truth no matter what the interpretation. To see that, under \mathfrak{U}_1 , ϕ_1 is not metaphysically necessary, note that $\Box \phi_1$ is not true under \mathfrak{U}_1 . Therefore, $\sim \Box \phi_1$ is true under \mathfrak{U}_1 . That is, under \mathfrak{U}_1 , ϕ_1 is not metaphysically necessary.

Claim 4: Under \mathfrak{A}_1 , ϕ_1 is an analytic truth which is not necessary.

To see this, focus on what ϕ_1 says. If one considers just what the logical constants 'if-then', 'the', and 'there is' mean, it seems reasonable to suggest that ϕ_1 is true solely in virtue of the meanings of these words. Since in this interpretation, ϕ_1 is not necessary, we have an example of an analytic truth that is not necessary. If one just grants that the logical truths form a subclass of the analytic truths, then this claim follows immediately from the others.

It turns out that we need not have relied on the semantics of rigid descriptions to find sentences for which claims like the above hold. We could have used a simple modal predicate logic that has both nonrigid descriptions and an operator on formulas which, without regard for modal contexts, looks back to the base world to evaluate the truth of the formula. So, for example, in a modal logic with an actuality operator \mathcal{A} , we can still generate LTNNs and ATNNs. To verify this, suppose we added such an operator to our language by generating formulas of the form $\mathcal{A}\phi$ in the clause for complex formulas. To interpret \mathcal{A} , we need the following clause for satisfaction

of formulas of the form $\mathcal{A}\phi$: f satisfies $\mathcal{A}\phi$ at w iff f satisfies ϕ at w₀. This guarantees that $\mathcal{A}\phi$ is true at a world iff ϕ is true at the actual world. Furthermore, let descriptions of the form $(\iota x)\psi$ be interpreted nonrigidly. Now consider the following formula:

$$(\phi_2) P(\iota x) \mathcal{A} Q x \rightarrow (\exists y) Q y$$

This is to be read: If the actual Q-thing is P, then something is Q. Such a sentence is a logical truth, since at the actual world of every interpretation, the truth of the antecedent materially implies the truth of the consequent. But, under an interpretation like \mathfrak{A}_1 (described above), ϕ_2 is not necessary. The reader should verify that claims similar to Claims 1-4 hold with respect to this conditional as well.

In fact, an even simpler example can be constructed in the modal *propositional* calculus with an actuality operator. Consider the following formula schema:

$$(\phi_3)$$
 $\mathcal{A}\psi \rightarrow \psi$

Any instance of this schema in the modal propositional calculus will be logically true, since, again, in any interpretation, the antecedent materially implies the consequent. But, under an interpretation where ψ is true at w_0 and false at some other world, such an instance of ϕ_3 is not necessary. The reader should verify that claims similar to Claims 1-4 above are true with respect to instances of ϕ_3 .

So rigid descriptions are not the only means of distinguishing logical truth, analytic truth, and metaphysical necessity. One should beware of the inference from logical truth to metaphysical necessity even in languages without rigid descriptions, for there may be an (implicit) actuality operator somewhere that invalidates the move. It is surprising that, by merely adding expressive power that allows us to talk about what is actually the case, the modal propositional calculus involves distinctions that undermine the traditional view that logical truths are necessary.

II

If correct, these results call for a reassessment of the traditional understanding of the notion of logical truth and its relation to metaphysical necessity. There are, however, ways one might try to challenge the claims in the previous section. One way would be to argue that we have used the wrong definition of logical truth. Recently, an alternative to the definition of logical truth used in section I has found its way into the literature. It is a consequence of this alterna-

⁵ See J. Crossley and I. Humberstone, "The Logic of 'Actually'," Reports on Mathematical Logic, VIII (1977): 11–29.

tive definition that ϕ_1 - ϕ_3 are not logical truths and, hence, not LTNNs. But, without begging any questions, it can be shown that this alternative definition of logical truth is incorrect.

Note, first of all, that our definition of logical truth is the standard Tarskian-Kripkean notion of truth-in-all-interpretations as it applies to our particular modal language. Logical truth is defined in terms of truth under an interpretation, and truth under an interpretation is defined as truth under an interpretation with respect to the actual world. This clearly squares with Kripke's original definition of validity. Recall that Kripke defined models φ that recursively assigned each formula a truth value relative to each world on a given model structure. Valid formulas are those which receive the assignment T at the actual world for every model on a model structure ("Semantical Considerations," p. 64; in Kripke's notation, a formula A is valid iff $\varphi(A, G) = T$, for every model φ on a model structure (G, K, R), where G is the real world, K is the set of all worlds, and R is the accessibility relation). The only real difference is that, for simplicity, we have ignored the accessibility relation.

Recently, certain modal and intensional logicians have deviated from Kripke's definitions, in a seemingly innocuous way, by dropping the distinguished world from models. Instead of the above series of definitions, they proceed directly from the definition of truth-under-an-interpretation-with-respect-to-a-world to the definition of logical truth, without the intermediate definition of truth under an interpretation. Theorists such as G. E. Hughes and M. J. Cresswell, Richard Montague, E. J. Lemmon and Dana Scott, 8 Brian Chellas, D. Dowty, F. Wall, and S. Peters, 10 and J. F. K. van Benthem, 11 to mention a few, use a definition of " ϕ is true under $\mathfrak A$ at world w" similar to the one above, but then directly define " ϕ is logically true (valid)" as: $(\forall \mathfrak{A})(\forall w)(\phi)$ is true under \mathfrak{A} at w). These authors never say why they alter Kripke's definition. The reason may have stemmed from their natural zeal to avoid metaphysical questions while doing logic. By eliminating the actual world from the semantics of modal logic, one can avoid the question of what it is for

⁸ An Introduction to Modal Logic, American Philosophical Quarterly Monograph Series, XI (Oxford: Basil Blackwell, 1977), p. 24.

Modal Logic: An Introduction (New York: Cambridge, 1980),

⁶ An Introduction to Modal Logic (London: Methuen, 1974), pp. 71, 351. 7 "Universal Grammar," in Montague, Formal Philosophy, Richmond Thomason, ed. (New Haven: Yale, 1970), p. 229.

¹⁰ An Introduction to Montague Semantics (Boston: Reidel, 1981), p. 127. 11 A Manual of Intensional Logic, CSLI Lecture Notes (Number 1), Center for the Study of Language and Information, Stanford University, 1985, pp. 13/4.

a world to be actual, a question figuring prominently in the literature. Since the alternative definition seems to be equivalent to Kripke's, it therefore seems more attractive. Hughes and Cresswell say that the definition of validity as truth in every model in Kripke's sense is "precisely equivalent" to its definition as truth in every world in every model (351). The idea seems to be that the permutation of the base world with any other world in a Kripkean interpretation produces a perfectly good new Kripkean interpretation, and so, ultimately, every world in such interpretation will play the role of the base world in some interpretation or other. So it seems that the alternative definition makes the arbitrariness of the choice of the base world explicit. Dowty, Wall, and Peters say as much in their recent text (op. cit.):

Thus, the definitions themselves do not single out any particular world as the actual ones; we may alternatively choose one or the other of them as the "actual" world, all other worlds becoming possible but not actual worlds relative to this choice. (Alternatively, we could change the definition of a model to include the designation of one particular world as the actual one—in fact, Kripke's original treatment followed this procedure. But with his approach there will be alternative models differing only in the choice of which world is so designated. Thus the definition of validity in our approach as truth in every world in every model is equivalent to the definition of validity in Kripke's approach as truth in every model) (127).

The problem is, however, that the two definitions are equivalent only with respect to modal and intensional languages of the kind these authors describe-languages that do not have rigid descriptions or actuality operators. To see that the alternative definition of validity is not always equivalent to Kripke's, consider our language and note that, on their definition, ϕ_1 fails to be a logical truth. It fails to be true in \mathfrak{A}_1 , for example, since an arbitrarily chosen assignment f satisfies P(1x)Qx at world W_1 (the object that is Q in the actual world is P in every world), but fails to satisfy $(\exists y)Qy$ at W_1 (since nothing is in the extension of Q at w_1). So there is an interpretation and world where ϕ_1 fails to be true, and this shows that ϕ_1 fails to be logically true on the alternative definition. Hence, the two definitions are not equivalent with respect to our language. None of the above authors acknowledge or hint that the two definitions are not equivalent for languages with rigid descriptions or actuality operators (but, then, none develop a language that contains such expressions). Nevertheless, a simple enrichment of the modal predicate calculus distinguishes the two definitions.

Which definition of logical truth is correct? If Kripke's is, our work in section 1 must be taken seriously; if the alternative is, it need not be. Is there an argument that shows the alternative definition to be correct, other than the question-begging argument which so concludes on the grounds that there are no LTNNs? Alternatively, would we be begging the question in the opposite direction if we were to claim that the alternative definition is a conflation of the formal notion of validity (truth-under-all-interpretations) with the metaphysical notion of necessity (truth-at-all-possible-worlds)? The answer to this last question is No!, and the reason is that there are independent grounds for claiming that the alternative definition is incorrect. Those grounds are just the following: (1) the most important semantic definition for a language is the definition of truth under an interpretation, and the alternative method, in which no world is distinguished as the actual world, has no means of defining this notion; and (2) the semantic notion of logical truth is properly defined in terms of the semantic notion of truth, and the alternative definition of logical truth is the wrong one because it fails to do this.

Once we take Tarski's notion of truth seriously, it becomes immediately clear why Kripke distinguished one of the worlds in each model. The semantic notion of truth-under-an-interpretation for a modal language with Kripke-style semantics requires a distinguished world, as a matter of logic! That is, the very logical setup requires a distinguished world. And Tarski's insight is that logical truth is truth-under-all-interpretations. When we define a modal language, we need a definition of truth-under-an-interpretation if we are to define logical truth by applying Tarski's insight. Kripke's models permit the definition of logical truth to follow this pattern (the alternative models do not), and when we enrich modal languages interpreted by such models, by adding rigid descriptions or actuality operators, the definitions of truth and validity remain essentially the same. So we do not beg any questions when we argue that the alternative definition of logical truth is incorrect. It fails to take the notion of truth seriously. A second look at the alternative definition reveals it to be a conflation of the semantic notion of validity and the metaphysical notion of necessity.12

¹² This conclusion is much stronger than those in Martin Davies and Lloyd Humberstone, "Two Notions of Necessity," *Philosophical Studies*, XXXVIII, 1 (July 1980): 1–30. These authors construct new technical distinctions based on what they see as just two different notions of validity—they call the alternative definition of validity "general" validity and Kripke's original notion of validity "real world" validity. So ϕ_1 – ϕ_3 are real-world valid, though not generally valid. They employ a

A second way to challenge our claims in section I would be to argue that the notions "the" and "actually," represented by the term-forming operator '1' and the sentence-forming operator ' \mathcal{A} ', respectively, are nonlogical notions. If this could be established, then $\phi_1 - \phi_3$ would not be logical truths and, hence, not LTNNs. Ideally, the question of whether or not these notions are logical should be decided by appeal to a clear definition of "logical notion." Unfortunately, there is no definition that identifies all and only the genuine logical notions. Nor do we plan to try to produce such a definition in the present paper. Consequently, we cannot say with certainty that the notions represented by our operators are logical notions. Nevertheless, on the traditional understanding of what is logical and what is not, our operators and the notions they represent seem to be logical.

Part of the traditional conception of a logical operator is this: the logical operators of a language are the operators evaluated in the recursive clause of the definition of truth. Clearly, our actuality operator is one of these. In addition, in the definition of the language itself, the complex formulas that involve this operator are generated in the same clause that generates the other logically complex formulas. The traditional conception of logic also regards the definite description operator as logical. Traditionally, it has been defined in purely logical terms. Moreover, the argument which concludes that the notions represented by our operators are nonlogical on the grounds that they are semantically interpreted by reference to the actual world, is not a good argument. We have just seen that, as far as the semantics of modality goes, an actual world must be distinguished as a matter of logic. This is essential to the definition of

new operator \mathcal{F} ("fixedly"), where " \mathcal{F} ϕ " is true just in case ϕ is true in all interpretations that differ only with respect to which world is designated as actual. And, following Gareth Evans ["Reference and Contingency," *The Monist*, LXII, 2 (April 1979): 161-189], they distinguish "deep" necessity (\mathcal{F} \mathcal{A} ϕ) from "superficial" necessity (\mathcal{F} \mathcal{A} ϕ).

If one accepts the alternative definition of validity, then these distinctions can certainly be drawn, even if they are not supported by any direct evidence. But the notion of general validity does not seem to be a proper notion of logical truth, since it is not defined in terms of the notion of truth. It appears to conflate the notions of logical truth and necessity. Moreover, Davies and Humberstone do not make it clear that the most important semantic notion, namely, truth simpliciter, cannot be defined in modal languages without a world distinguished as actual. Consequently, there seems to be no reason to distinguish deep necessity and superficial necessity in order to try to preserve the intuition that logical truths are necessary. It seems much more realistic to admit that this intuition is false.

truth. So the fact that our operators exploit this distinction does not

imply that they are nonlogical.

Consequently, the suggestion that our operators are nonlogical appears to conflict with the traditional understanding of what is logical. Of course, the traditional conception may be mistaken, but, if so, our claims in section 1 might then become part of the argument to show that the traditional conception rests on a mistake.

We have not found in the literature any examples of LTNNs and ATNNs simpler than those described in section 1.13 Even though Kripke stipulates that analytic truths are necessary, one might suspect that, somewhere in his discussion in "Naming and Necessity," there are bound to be some simple LTNNs or ATNNs. He does not appear, however, to regard typical English descriptions as rigid designators. Consequently, his examples in which definite descriptions are used to fix the reference of certain terms could not be construed as LTNNs or ATNNs.14 The other prominent examples of Kripke's lectures cannot be employed in the demonstration that there are LTNNs either. The examples we have in mind here are the following:

- (1) The inventor of bifocals invented bifocals.
- (2) The inventor of bifocals might not have invented bifocals.
- (3) The teacher of Alexander taught Alexander.
- (4) The teacher of Alexander might not have taught Alexander.

Let us assume that the definite descriptions that constitute the subject terms of these sentences are genuine, ineliminable logical units. For the moment, let us not be concerned with the Russellian analyses in which the descriptions are eliminated, since these do not do justice to their apparent subject-predicate form. The logical form of (1) and (3) seems to be captured by (5); the form of (2) and (4) is captured by (6):

- (5) P(1x)Px
- (6) $[\lambda y \lozenge \sim P y](1x) P x$

We may read 'Px' as either "x invented bifocals" or "x taught Alexander." Now, if (1) and (3) are analyzed as (5), then they seem to be true. If (2) and (4) are analyzed as (6) (where the description is still

¹³ Davies and Humberstone (op. cit.) informally describe a variant of ϕ_2 . 14 For example, by fixing the reference of 'Neptune' with the description 'the planet causing certain perturbations', we do not get LTNNs by biconditionalizing "Neptune exists" with "Some unique planet causing certain perturbations exists," unless we replace 'Neptune' in the resulting biconditional with a rigid description; see Kripke, "Naming and Necessity," pp. 347/8, note 33.

rigid), then they also seem to be true. Moreover, one might think that (1) and (3) are LTNNs. For it looks as if what (5) says is logically true, and it looks as if we can use λ -conversion on (6) to show that (5) is not necessary (λ -conversion is a valid rule when used with rigid descriptions).

But this is a mistake. Sentences like (5) are not logically true, since there are many interpretations in which the description fails to denote. And, when it does, such sentences are not true, by the definition of satisfaction. Consequently, (1) and (3) are not LTNNs, since they are not examples of logical truths. Nor are they ATNNs. They are not analytic, because their truth does not depend solely on the meanings of their words and their arrangement (their truth depends on the fact that something satisfies the description). So these examples cannot be used to establish the claims in this paper. One should note that (2) and (4) are English sentences which are most plausibly analyzed by construing the description as rigid.15 A nonrigid description would not help us capture the truth expressed by (2). Nor would these examples be of any help if we were to use the Russellian elimination of the descriptions. It is a straightforward exercise to show that none of the various Russellian readings of (1)-(4) are either LTNNs or ATNNs.

There are in the literature some extraordinary examples of LTNNs, but these do not identify the point at which the distinction between logical truth and necessity first evolves. In his work on demonstratives, David Kaplan offers "I am here now" and "I exist." The idea is that these sentences, under any possible interpretation and in any context of utterance, will always be assigned as content a proposition which is in fact true but which might have been false. Thus, we have examples of LTNNs. There are, however, several reasons why these examples do not cut to the heart of the distinctions we are after.

For one thing, Kaplan's examples require a much more powerful logical apparatus than simple (quantified) modal logic. They involve the indexical 'I' and are expressed in the language that forms the basis of the logic of demonstratives. This logic involves indexicals,

(February 1978): 81–98; also published in French, Uehling, and Wettstein, op. cit., pp. 401–412.

¹⁵ See Baruch Brody, "Kripke on Proper Names," pp. 75–80, in *Contemporary Perspectives in the Philosophy of Language*, French, Uehling, and Wettstein, eds. (Minnesota UP, 1979), pp. 75–80.

p. 73 (page reference is to the unpublished manuscript, Draft # 2).

demonstratives, tense operators, contexts of utterance, agents, positions, times, and the distinction between character and content. Kaplan even suggests that the distinction between character and content is the key to distinguishing logical truth and necessary truth:

The bearers of logical truth and of contingency are different entities. It is the *character* [his emphasis] (or, the sentence, if you prefer) that is logically true, producing a true content in every context. But it is the content (the proposition, if you will) that is contingent or necessary ("Demonstratives," 72).

There is something to what Kaplan says here, but we have now seen examples of this phenomenon in a logic which is far simpler than the logic of demonstratives. The distinction between character and content, and the rest of the apparatus involved in the logic of demonstratives, is not required to construct the definitions that are essential for separating the logically true from the necessary.

There are two other reasons why Kaplan's examples do not illuminate the basic insights about the relevant notions. One is that, if the examples are to work, we must accept a proper thesis of metaphysics. The required thesis is: either there are no nonexistent objects or, if there are, they cannot be the agents of real utterances. For if there are nonexistent objects in the domain of individuals and they are allowed to be the agents of utterances, then there will be interpretations and contexts in which the indexical 'I' denotes a nonexistent object. In such interpretations and contexts, the sentences "I am here now" and "I exist" will be false. Hence, they will not be logical truths. Recent work in philosophy reveals that the thesis that there are nonexistent objects is viable. 18 The present point, however, does not depend on one's inclination (not) to believe in nonexistent objects or on the natural hypothesis that they cannot be the agents of ordinary contexts of utterance (they can be agents in fictional contexts). The point simply is that proper metaphysical theses are required for these examples to be LTNNs. Our LTNNs do not require such nonlogical theses.

One final reservation concerning these examples is that it is unclear whether they are ATNNs. Kaplan asserts that one need only understand the meaning of "I am here now" to know that it cannot be uttered falsely. Even if one agrees with Kaplan about this, it is

19 "On the Logic of Demonstratives," op. cit., journal version,

¹⁸ See Terence Parsons, "A Prolegomena to Meinongian Semantics," this JOUR-NAL, LXXI, 16 (Sept. 19, 1974): 561–580; Nonexistent Objects (New Haven: Yale, 1980); and Zalta, op. cit.

unclear whether this still counts as an analytic truth, since contexts of utterance are now involved. That is, this is an example of an analytic truth only if one uses a somewhat expanded notion of analyticity. We cannot consider the truth of the sentence without appealing to some context, and so we cannot simply say that it has the property that traditional analytic truths have, namely, being true in virtue of the meanings of its words. Rather, it has the property of being true in all contexts in virtue of the meanings of its words relative to such contexts. Although the sentence seems to qualify as analytic in this expanded sense, the examples we have described are analytic in the stricter, traditional sense.

But Kaplan has another example which is more instructive than the previous two. It is expressed with his operator 'dthat', which turns any term that it prefaces into a rigid designator. 20 In Kaplan's system, a definite description such as $(\iota x)\phi$ is not rigid, but ' $dthat[(\iota x)\phi]$ ' is (we assume some familiarity with the 'dthat' operator). The example Kaplan offers is: $\tau = dthat[\tau]$, where τ is a definite description ("Demonstratives," 71). The idea here is that no matter what interpretation and assignment to the variables one considers, the assignment satisfies this sentence with respect to the actual world because the denotations of the two descriptions are identical at that world. The sentence is not necessary, however, because there will be worlds in which the denotations of the descriptions differ. Although this example is also expressed in the language underlying the logic of demonstratives, only a very small portion of the language and logic of demonstratives is required for its analysis. Even if we resect, from the logic of demonstratives, just the apparatus required for the analysis, our examples in section I get closer to the crux of the distinction between logical truth and necessary truth.

There are two reasons for this. The first is that they do not require a modal logic with identity formulas, and so do not require that we face, as Kaplan must, the host of vexing questions about the interactions of identities and modal contexts. The second, and more important, reason is that our examples do not require any special treatment of nondenoting descriptions. Kaplan's sentence works as an example of a logical truth that is not necessary only because he has an artificial way of treating descriptions that no object satisfies. Here is why.

In Kaplan's system, a description that no object satisfies is assigned an arbitrary, "alien" entity, and so descriptions never fail to de-

²⁰ See "Dthat," in French, Uehling, and Wettstein, op. cit., pp. 383-400.

Digitized by Arya Samaj Foundation Chennai and eGangotri

note.21 Without this device, there would be numerous interpretations in which both $(\iota x)\phi$ and $dthat[(\iota x)\phi]$ fail to denote anything whatsoever. In such interpretations, Kaplan's sentence would be false, for his truth definition, like ours, is constructed in accordance with the correspondence theory of truth. It guarantees that atomic and identity sentences are true iff both all of their terms denote and the denotations stand in the right relationship. Hence, such sentences would not be logical truths. Kaplan's technique of having descriptions that no object satisfies denote the alien entity does preserve his sentence as a logical truth—when nothing satisfies $(ix)\phi$, $(\iota x)\phi = dthat[(\iota x)\phi]$ is true, because both descriptions denote the alien entity. But this piece of logician's art undermines whatever insight this example might have offered those trying to track down the point at which logical truth and necessity first diverge. Nothing in the logic of demonstratives is needed for identifying LTNNs.

The traditional concepts of logical truth and metaphysical necessity are therefore independent, though not mutually exclusive. Identity statements involving rigid names and descriptions are examples of necessary truths that are not logically true, and the axioms of propositional logic are examples that are both logically true and necessarily true. Since the logical truths form a subclass of the analytic truths, it follows that the concepts of analytic truth and necessity are also independent, though not mutually exclusive. The same examples suffice to show this. One preliminary conclusion we can draw from the foregoing is that there is at least one sense of 'logical' for which the locutions 'logically possible' and 'logically necessary' represent confusions of the formal and material modes. For the statement that ϕ is logically possible, where ϕ is some English sentence, may represent a conflation of the formal statement that ϕ is satisfi-

the unique $i \in \mathcal{U}$ such that $\models_{c,f(\alpha/i),t,w}^{\mathcal{U}} \phi$, if there is such; the alien entity, otherwise

Essentially, the clause says that the denotation of '(the α) ϕ ' with respect to context C, assignment f, time t, and world w is the unique individual i such that ϕ is true relative to $C, f(\alpha/i), t, w$, where $f(\alpha/i)$ is the assignment just like f except that it assigns i to α . The problem with this clause is that it obscures the logical source of the uniqueness. Logically speaking, it is the uniqueness of a certain assignment function satisfying ϕ that guarantees that the description '(the α) ϕ ' has a denotation. And to assert that $f(\alpha/l)$ is unique, one must quantify over other possible assignment functions. But the clause Kaplan uses does not appear to do this. Compare our Clause 3 in section 1. I do not think this is just a matter of taste; rather, it is a matter of finding the logically correct way of formulating a semantic version of Russell's syntactic analysis.

²¹ There is a technical unclarity in the clause Kaplan uses to assign denotations to descriptions. Consider his clause in "On the Logic of Demonstratives," p. 90: if $\alpha \in \nu_i$, then $|(the \ \alpha)\phi|_{c,t,t,w}^u =$

Digitized by Arya Samaj Foundation Chennai and eGangotri able (true in some interpretation) with the material statement that ϕ , under interpretation we know it to have, is true in some metaphysically possible world. Typically, when philosophers say that ϕ is logically possible (and one does not go far in philosophy before encountering such statements), they mean the latter. But, if so, we have seen that this kind of possibility is independent of logic.

Our formal work clearly shows that there are contingent truths known a priori, regardless of whether anyone fixes the reference of proper names. The formulas ϕ_1 , ϕ_2 , and ϕ_3 are all examples of this. Since they are logical truths, they are all true and, hence, possible. Since none are necessary, their negations are possible as well. Given that they are possible and that their negations are possible, they must be contingent. Since they are analytic truths, they can be known a priori. The claim that there are ATNNs, then, is stonger than Kripke's claim that there are contingent truths known a priori, for the former implies the latter, but not vice versa.

We also ought to be aware that the definition of a valid argument as an argument in which it is impossible that the premises be true and the conclusion false, something we all learn/teach in a first philosophy class, is incorrect. The argument that has the antecedent of ϕ_2 as a premise and the consequent of ϕ_2 as a conclusion is surely valid. Yet it is one where it is metaphysically possible for the premise to be true and the conclusion false. It is always good to know the exact place where we have bent the truth a little to simplify a concept.

Finally, it may still come as a surprise that there are logical false-hoods that are metaphysically possible. The negations of ϕ_1 , ϕ_2 , and ϕ_3 are examples of this. Since ϕ_1 , ϕ_2 , and ϕ_3 are logically true, their negations are logically false. Since ϕ_1 , ϕ_2 , and ϕ_3 are not necessary, their negations are metaphysically possible. Hence, their negations are logical falsehoods that are metaphysically possible. Contemplation of these examples should rid us of the automatic tendency to regard the self-contradictory quality a modal statement may have as the justification for why it could not be the case.²²

EDWARD N. ZALTA

Stanford University

²² Subsequent to the submission of this paper, Joseph Almog's paper, "Naming without Necessity," was published in this JOURNAL, LXXXIII, 4 (April 1986): 210–242. In the last section of the paper, Almog identifies several contingent propositions that are "structurally true." It is unclear what the relationship is between structural truth and logical truth, since the notion of a structural truth is not defined. But, more importantly, Almog's examples count as contingent structural truths only because he has adopted essentialist and actualist metaphysical assumptions. Examples (3), (4), and (5) (240/1) require essentialist assumptions to

Digitized by Arya Samai Foundation Chennal and eGangotri APPENDIX

We provide the reader with an example to show that the definitions of denotation and satisfaction work correctly. For simplicity, we ignore the relativization to possible worlds. Suppose f satisfies P(1x)Qx. Then (30)(0 = $d_{\mathfrak{A},\mathfrak{f}}((1x)Qx)$ & $o \in F(P)$). Call an arbitrary such o, " o_1 ." So o_1 = $d_{\mathfrak{A},\mathfrak{f}}((1x)Qx)$ & $o_1 \in F(P)$). By the definition of denotation, the left conjunct is equivalent to: $(\exists!f')(f' = f \& f'(x) = o_1 \& f' \text{ satisfies } Qx)$. Call the unique such "f, f_1 ." We now have, $f_1 = f \& f_1(x) = o_1 \& f_1$ satisfies $Qx \& f_1$ $o_1 \in F(P)$. By the definition of satisfaction, the third conjunct is equivalent to: $(\exists o')(o' = d_{\mathfrak{A},i_1}(x) \& o' \in F(Q))$. If we call such an o', " o_2 ," we know: o_2 = $d_{\mathfrak{A},\mathfrak{t}_1}(x) \& o_2 \in F(Q)$. But, since the denotation of a variable is just what the assignment assigns to it, we have: $O_2 = O_1$. Therefore, the third conjunct becomes equivalent to: $O_1 \in F(Q)$. Reassembling our four part conjunction, we get: $f_1 \stackrel{\times}{=} f \& f_1(x) = o_1 \& o_1 \in F(Q) \& o_1 \in F(P)$. In other words, f satisfies P(ix)Qx iff $(\exists!f')(\exists o)(f' \stackrel{x}{=} f \& f'(x) = o \& o \in F(Q) \& o \in F(P))$. These conditions are the correct satisfaction conditions, and they are ones in which the notions of denotations and satisfaction do not appear.

qualify as contingent, and example (6) (241) requires actualist assumptions to qualify as structurally true. In contrast to this, our examples require no proper thesis of metaphysics to count as LTNNs and ATNNs—they cut directly to the origin of the distinctions among logical, analytic, and necessary truth.

Digitized by Arya Samaj Foundation Chennai and eGangotri SECOND-ORDER LOGIC STILL WILD*

N Philosophy of Logic, W. V. Quine summed up a popular opinion among mathematical logicians by referring to second-order logic as "set theory in sheep's clothing." A few years later George Boolos argued that Quine's claims were overblown. He conceded that second-order logic commits one to some classes—every subclass of one's individual domain, to be exact—but he also observed that in most cases this falls far short of the commitments of the usual set theories.

I.

In two very interesting recent papers,² Boolos has tried to tame second-order logic completely. Second-order logic *per se*, he now claims, does not commit one to sets, classes, Fregean concepts, or anything else. The gist of his argument is this:

- 1. We need not posit classes or collections in order to render second-order sentences intelligible. We can simply translate them into ordinary language using plural quantifiers. (For example, the second-order version of the least-number principle, "For any F, if there are numbers that F, then there is a least such" may be translated as "It is false that there are some numbers such that no one of them is the least.")
- 2. Using plural quantifiers does not commit one to classes or collections. Indeed, it does not commit one to anything that one is not already committed to by means of one's use of singular quantifiers.
- 3. Thus, the use of second-order logic need not commit one to collections or sets. Quine is wrong; second-order logic is not class theory in disguise.

If this is correct, Boolos has made a significant contribution to the philosophy of mathematics. He will have vindicated those who have held that conflating second-order logic with set theory has perverted Frege's logicism, and he will have strengthened their version of logi-

Englewood Cliffs, NJ: Prentice-Hall, 1970; reprinted, Cambridge, MA: Har-

vard, 1986.

^{*} I am grateful to Susan Hale, Bill Lycan, Mark Risjord, Jay Rosenberg, Geoffrey Sayre-McCord, Stewart Shapiro, and Steven Wagner for helpful comments.

² Boolos's initial criticism of Quine occurs in his "On Second-order Logic," this JOURNAL, LXXII, 16 (Sept. 18, 1975): 509–537. His two recent papers are "To Be Is to Be a Value of a Variable (or to Be Some Values of Some Variable)," this JOURNAL, LXXXI, 8 (August 1984): 430–449; and "Nominalist Platonism," *Philosophical Review*, XCIV, 3 (July 1985): 327–344. I will refer to these as TB and NP, respectively. The bulk of the argument that I here attribute to Boolos may be found in TB.

Digitized by Arya Samaj Foundation Chennai and eGangotri pitfall he fell cism. He will also have helped Hartry Field out of a pitfall he fell into while trying to nominalize Newtonian mechanics.4 He will have shown structuralists, such as Stewart Shapiro and me, how we can use second-order characterizations of set-theoretic hierarchies without fear of circularity.5 Finally, he will have shown set theorists how they can embrace second-order axioms without abandoning the idea that the first-order universe of sets contains all the setlike entities that there are. (Shapiro tells me that this has been the principal motive for Boolos's work on second-order logic.)

Although one might see Boolos's translation as a "nominalization" of second-order logic, it does not significantly affect the central ontological problem in the philosophy of mathematics. Unless we adjoin it to a program such as Field's, which has other problems besides its use of second-order logic, Boolos's methods do not dispense with sets and other abstract objects. At most, his technique eliminates the use of abstract entities from several specific (but important) contexts.

Despite the attractiveness of Boolos's view for philosophers of mathematics, and for me in particular, I am not convinced that it is an advance over the standard view of second-order logic. Here are my reservations. First, my logical/linguistic intuitions differ sharply from Boolos's. As I see it, some plural quantifications do commit us

³ Second-order logic has many friends, but, for someone who wants to combine it with logicism, see J. M. B. Moss, "The Mathematical Philosophy of Charles Parsons," British Journal for the Philosophy of Science, XXXVI, 4 (December 1985): 437-455, p. 447; and his review of my Frege and the Philosophy of Mathematics, this JOURNAL, LXXIX, 9 (September 1982): 497-512.

⁴ Field abandoned his use of second-order logic in "On Conservativeness and Incompleteness," this JOURNAL, LXXXII, 5 (May 1985): 239-260. During a commentary at the 1984 Eastern Division meeting of the American Philosophical Association, however, he stated that Boolos's work had convinced him that he could use second-order logic to achieve a nominalist formulation of Newtonian physics over which Platonist mathematics is conservative.

⁵ The structuralist's dilemma is this. We maintain that mathematics is a science of patterns or structures and that sets are positions in a particular type of structure or structures—the (an) iterative hierarchy of sets. However, the fullest accounts of such structures use second-order logic. Thus, if we construe second-order logic as quantifying over setlike entities, we seem to be presupposing the notion of set to say what a set is. Of course, this is an old difficulty and not unique to the structuralist approach. And there are ways out: (1) One can maintain that the second-order variables range over sets qua manys rather than sets qua ones or over Fregean concepts, or over properties, where each of these are taken to be logical entities rather than elements of a structure. (2) Or one can claim that we just cannot hope to get fully complete and categorical descriptions of the richer mathematical structures. (This is the position toward which I now incline.) Boolos's approach would wipe out the problem of circularity in one quick blow. For further discussion, see Stewart Shapiro, "Second-order Languages and Mathematical Practice," Journal of Symbolic Logic, L, 3 (September 1985): 714-742.

to collections, sets, or some other sort of second-order entity, and I find fault with Boolos's arguments against my position. I will expand on this in section II. But there is more to deciding matters of ontic commitment than brandishing logical and linguistic intuitions. Boolos may have had this in mind when he constructed an alternative semantics for second-order logic according to which no secondorder variable has a second-order entity as one of its values. In section III, I argue that, although this semantics is necessary for the consistency of Boolos's position, it fails to demonstrate that secondorder logic is not committed to collections. In section IV, I argue that neither the intelligibility of plural quantifications nor their prima facie lack of commitment to collections is sufficient to demonstrate that they never commit us to collections. Determining whether they do involves representing them in an acceptable and suitably interpreted logical notation. Thus, those of us who find plural quantification in need of logical analysis will not be enlightened by formal explications such as Boolos's, which presuppose plural quantification for their interpretation.

II.

I will start with my dissenting logical/linguistic intuitions. I find that I cannot process many sentences containing plural quantifications without understanding them in terms of collections. Take one of Boolos's leading examples, the Geach-Kaplan sentence,

(1) Some critics admire only one another.

On the face of it this makes no mention of classes or collections. However, I am inclined to understand it as saying:

(2) There is a nonempty collection of critics each member of which admires no one but another member.

Boolos urges me to see (1) as a second-order sentence, symbolized by

(3) $(\exists X)[(\exists y)Xy \& (x)(y)(Xx \& Axy \rightarrow x \neq y \& Xy)]$

and to translate this, using his general scheme, into English as

(4) There are some critics such that any one of them admires another critic only if the latter is one of them distinct from the former.

But this sentence seems to me to refer to collections quite explicitly. How else are we to understand the phrase 'one of them' other than as referring to some collection and as saying that the referent of 'one' belongs to it?

Of course, if we render (3) and (4) as (1), we can avoid the problematic phrase 'one of them', and see (4) as just an awkward way of putting (1). But I have two worries with this way out. First, we have no assurance that we can always smooth out Boolos's translations in this way. Second, I started out by worrying about the ontic commitment of (1). The buck was passed to (3) and thence to (4). If (4) is just an awkward way of putting (1), we have come full circle.

Perhaps, I have been looking in the wrong place for a reference for 'them' in 'one of them'. Could not Boolos reply that plural quantifiers and the plural pronouns they govern do not refer singly to collections but rather divide their reference over the members of collections? I think he could. The notion of divided reference makes good sense with respect to predicates: 'sleek' refers in the divided sense to a locomotive just in case "sleek" is true of it. Of course, neither quantifiers nor variables refer to specific individuals outright. We might, however, extend the notion of reference to quantifiers and variables as follows: a quantifier, with its associated variable, refers in a given sense just in case it generalizes over positions that are open to terms that refer in that sense. Then second-order quantifiers, by virtue of generalizing over predicate positions, would count as referring in the divided sense. Of course, although this fledgling theory of divided reference helps Boolos's case, it does not show that my intuitions concerning 'one of them' are wrong.

But there is more: Boolos argues that it is often a mistake to represent plural quantifications as referring to collections. He claims that

(5) There are some sets that are self-identical, and every set that is not a member of itself is one of them.

should not be rendered as

(6) There is a collection of sets that are self-identical, and every set that is not a member of itself is a member of this collection.

For, he claims, the former is obviously true but the latter is false (TB, 446/7). Note, however, that (6) need not be false provided collections are ultimate classes or something similar. So we could put Boolos's point in more neutral terms as follows: (5) is obviously true but (6) is not; thus one should not paraphrase (5) as (6).

- Is (5) obviously true? Yes, if we read it as
- (7) There are some sets that are self-identical, and every set that is not a member of itself is a set.

For this is paraphrasable as the first-order sentence:

(8) $(\exists x)(Sx \& x = x) \& (x)(Sx \& x \text{ is not a member of } x \to Sx).$

Should we read (5) as (7)? Boolos thinks that the answer is an unqualified affirmative. In just a paragraph beyond the one last cited we read:

. . . I am loath to abandon [my strong intuition] about the meaning of English sentences of the form "There are some As of which every B is one," viz. that any sentence of this form means the same thing as the corresponding sentence of the form "There are some As and every B is an A" (TB, 447).

From which Boolos concludes that (5) "is simply synonymous with" (7). Furthermore, in NP (331/2) he says that "There are some classes such that every infinite class is one of them" is only an awkward way of saying "There are some classes and every infinite class is a class," rather than a way of saying that there is a class (collection) of infinite classes. Again, taking this reading, we have no reason to use a second-order translation and arrive at a first-order triviality of ZF rather than a well-known ZF falsehood.

More generally, Boolos equates

- (9) There are some As such that every B is one of them.
- (10) There are some As and every B is one.
- (11) There are some As and every B is an A.

which are equivalent within second-order logic, if rendered as

- $(12) \cdot (\exists X) [(\exists x) Xx & (x)(Xx \rightarrow Ax) & (x)(Bx \rightarrow Xx)]$
- $(13) \ (\exists X)[(\exists x)Xx \ \& \ (x)(Xx \rightarrow Ax) \ \& \ (x)(Bx \rightarrow Ax)]$
- (14) $(\exists x)Ax & (x)(Bx \rightarrow Ax)$

But this surely should not count as substantial evidence that (9), (10), and (11) are "synonymous" forms. It is often possible to prove within a theory that some of its sentences are equivalent to sentences in one of its more elementary subtheories. Thus, it is well known that within counting theory we can prove that

- (15) There are three wise men.
- (16) There is a number that counts the wise men and equals 3.

are equivalent, and in this way reduce some counting theorems to (obvious?) truths of first-order logic with identity. However, to my knowledge, no one today wants to conclude that (15) and (16) are synonymous. Furthermore, it is generally agreed that (16) but not (15) commits one to numbers. By parity of reasoning, the contrast

⁶ Similarly, within set theory "nothing fails to be self-identical" is equivalent to "there is a set of all non-self-identical things and it has no members"; yet these are not counted as synonymous.

Digitized by Arya Samaj Foundation Chennai and eGangotri between the apparent commitment of (9) and (10) to collections and (11)'s lack of it (as well as its logical transparency) should count as strong evidence that they are not synonymous.

I do not think we should grant Boolos his examples, because they depend upon the second-order equivalence of (12), (13), and (14), and the status of second-order logic is at issue. But suppose that we grant them anyway. Then they show that some plural quantifications should not be construed as quantifying over collections. But this is because these examples can be plausibly put into first-order terms. Thus, the examples do not bear on the question of whether what one might call "genuine" plural quantifications, such as,

(17) It is false that there are some sets such that every one of them has a member which is also one of them.

are wrongly construed as quantifying over collections. In (17) we have an example—Boolos's translation of the second-order axiom of foundation-which cannot be put in first-order terms. It is a substantial set-theoretic axiom and by no stretch of the imagination a triviality. Until Boolos can show us that construing it as quantifying over collections renders it plainly false or more controversial than it already is, my intuitions and his concerning it and the Geach-Kaplan sentence are at a standoff.

Since consulting intuitions concerning ordinary language has failed to resolve the issue between Boolos and me, it might be useful to turn to formal truth theories for second-order languages. If these require second-order variables to range over collections, then Boolos's translation would seem to show nothing. "Nominalist Platonism," Boolos's second paper, takes this possibility to heart. Truth theories now come to the forefront. With Quine's "To be is to be a value of a variable" in mind, Boolos writes:

The heart of the matter is this: it is only with respect to a truth-"definition" of the standard sort, a Tarski-style truth-theory for a first-order language, that the notion of a value of a variable is defined. In the case of a second-order language, such as the second-order language of set theory, there are at least two different sorts of truth-theory that can be given: on one of these, it would be quite natural to define "value" so that the second-order variables turn out to have classes as values; on the other it would not (NP, 324).

He then introduces a new truth theory which assigns no classes as values to second-order variables. The key idea is this: Instead of saying

Sequence s (which assigns individuals to the individual variables and classes of them to the predicate variables) satisifes "Xy" if and only if s(y) belongs to s(X).

Boolos says

Sequence s (which assigns values to only the individual variables) and relation R satisfy "Xy" if and only if R(X,s(y)).

Lapsing into class talk for a moment, we can see that $R\langle V,b\rangle$ holds if and only if (1) V is a second-order variable of the object language, (2) b is an element of the individual domain, and (3) b belongs to some class C(V,r), dependent on V and R, of members of that domain. Thus, given any sequence of the type used in a standard truth theory, i.e., one which assigns classes to the second-order variables, there is an equivalent relation R in Boolos's truth theory. Conversely, for any such R in a Boolos-style truth theory, there is an equivalent sequence of the usual type in an associated standard truth theory.

Notice that Boolos's relation R now takes the place of the problematic phrase 'one of them'. Moreover, we can even use R to give content to the idea that a predicate variable has a divided reference by stipulating that, with respect to s and R, V refers in the divided sense to b just in case $R\langle V,b\rangle$. (Cf. NP, p. 337.)

On this way of construing Quine's criterion, Boolos can certainly claim that, on his truth theory, classes are not values of variables. But I think that putting the issue of ontic commitment in terms of sequences and Tarski-style semantics obscures it. Sequences are something of an engineering trick, which allow one to have uniform recursion clauses in the truth definition for a quantificational language. Since, on the sequence approach, a first-order existential quantification is true only if an appropriate sequence exists, the ontic commitment to an individual that this entails is obscured. The question of the ontic commitment of second-order existential quantifications is even harder to evaluate in Boolos's truth theory. Because of this, I am going to present a variant of Boolos's truth theory which displays the ontic commitment of existential quantifiers (of both types) more perspicuously. I will start with the first-order case.

Definition. An interpretation I of a first-order language L in a nonempty domain D is an assignment of an element of D to each variable of L and an n-ary relation over D to each n-adic predicate letter of L. E^* is the assignment by I to the expression E.

Then we define truth for I in D as follows:

- (a) Fxyz is T for I in D iff $\langle x^*, y^*, z^* \rangle$ belongs to F^* .
- (b) $\sim S$ is T for I in D iff S is not T for I in D.

- (c) $S \vee W$ is T for I in D iff either S or W is T for I in D.
- (d) $(\exists v)S$ is T for I in D iff there is a d in D such that S is T for $I[v^*/d]$ in D.

Here $I[v^*/d]$ is the interpretation of L which is just like I except for assigning d (rather than v^*) to the variable v.

Clause (d) makes it clear that an existential quantification is true in a domain only if an appropriate element of the domain exists.

Let us try a similar approach to Boolos's truth theory. We cannot regard an interpretation as something that assigns entities to the second-order variables since Boolos carefully avoids doing that in his truth theory. Instead, let us regard an interpretation as a relation which relates variables and predicates to the appropriate elements of the domain. Of course, our interpretations for first-order languages are relations too, since all assignment functions are relations. But the relations that serve as Boolos-style interpretations need not be functions where the second-order variables are concerned.

More precisely, a Boolos-style interpretation I of a second-order language L in a nonempty domain D is a relation that relates each first-order variable v to a unique element v^* of D, each n-adic predicate letter F to a unique n-ary relation F^* over D, and each secondorder variable to 0 or more elements of D. Then we define truth for Iin D as follows:

- (a) Xy is T for I in D iff $\langle X, y^* \rangle$ belongs to I.
- (b) Fxyz is T for I in D iff $\langle x^*, y^*, z^* \rangle$ belongs to F^* .
- (c) -S is T for I in D iff S is not T for I in D.
- (d) $S \vee W$ is T for I in D iff either S or W is T for I in D.
- (e) $(\exists v)S$ is T for I in D iff there is a d in D such that S is T for I $[v^*/d]$ in D.
- (f) $(\exists V)S$ is T for I in D iff there is a relation R over D consisting solely of 0 or more ordered pairs of the form $\langle V, d \rangle$ (d in D) such that S is T for I[V/R] in D.

As before, $I[v^*/d]$ is the interpretation of L which is exactly like I except for assigning d to v. I[V/R] is the interpretation that is exactly like I except for containing the pairs in R in place of the pairs $\langle V, d \rangle$.

Comparing clauses (e) and (f) makes it quite clear that the truth conditions for the two types of existential quantifiers are parallel. Both use metalinguistic existential quantifiers. The only difference between them is that the condition for the second-order quantifier uses a second-order metalinguistic quantifier. Thus, one who is already convinced that second-order quantification commits one to

⁷ For simplicity I have restricted my account to languages without identity, constants, functional terms, or second-order variables beyond monadic ones. Also I have defined assignments over nonempty domains, whereas Boolos presented a truth theory for set theory, using the entire universe (a nonset) as his universe of discourse.

collections will find nothing to shake this conviction in Boolos's approach. To the firm believer this detour through Boolos's truth theory shows only that an *explicit* ontic commitment to classes need not emerge in a rigorous truth theory for a second-order language, but it fails to show that it is eliminable altogether.

Boolos admits as much. However, to those who are skeptical and to the "foes of classes" he offers the by now obvious way out: if you construe the second-order metalinguistic quantification used in his truth theories in terms of plural quantification, then you need not commit yourself to classes (NP, 337).

But now we are back to the issue of the last section. Whether we are a friend or foe of classes, we are likely to find it difficult to assess the ontic commitment of certain sentences containing plural quantifiers, such as the Geach-Kaplan sentence. Perhaps the most straightforward method for attacking this problem is to see it as a question of deciding whether such sentences should be represented in second-order logic or in class theory. But this suggestion—one which Boolos appears to endorse—will be of no avail unless we already know what ontic commitment second-order quantification carries. That is why telling us that it is the same as that of plural quantification hardly helps.

IV.

Boolos is involved in a circle: he uses second-order quantification to explain English plural quantification and uses this, in turn, to explain second-order quantification. Of course, the same is true of our standard explanations of English singular quantification in terms of first-order quantification and even of our account of English and truth-functional conjunction. We could not even begin the logical enterprise if we could not break some of these circles by taking some logical particles of ordinary language as reasonably clear starting points. Gottlob Frege was clearly aware of this. He objected to using ordinary language to settle logical matters, but he also argued that it is a suitable place at least to start to understand logical notations. I agree; to begin to understand the quantifiers, for instance, we need not have already grasped the Tarski semantics. Using ordinary language as a start, we can bootstrap ourselves into the language of logic and mathematics. In a related vein, Quine has emphasized the triviality of his criterion of ontic commitment: it simply reflects the idea that the existential quantifier explicates our ordinary concept of existence—that the quantifier ' $(\exists x)$ ' is just another way of saying "there exists an x."

⁸ These points deflect the charge that Quine's criterion is applicable only if it's unnecessary, because, in order to interpret the existential quantifier, one must first specify its domain and, thus, one's ontology. They also defuse the

Reflections such as these seem to be in the background when Boolos emphasizes that his scheme translates second-order sentences as *intelligible* English sentences which carry no explicit commitment to collections. Just as Quine finds 'there is' sufficiently clear to found his doctrine of ontic commitment, Boolos seems to think that the intelligibility of plural quantifiers shows that they are sufficiently clear to ground his doctrine of noncommitment. I take the opposite view.

Before I discuss Boolos's choice of circle, I want to remind us that, again at least since Frege, we have known that ordinary language is unsuitable for settling questions of logic and ontology. We have learned from examples, such as "Mary did it for John's sake" and "There is a possibility that it will rain today," that ordinary language often introduces excessive ontic commitments. Ironically, the Geach-Kaplan sentence is usually seen as a case in which ordinary language hides its ontic commitment.

It will be useful for us to consider hidden ontic commitment more carefully. Since Boolos has turned the Geach-Kaplan on its head,

here is another example for us to consider:

(18) Every time Saul visited Catherine he visited Matthew beforehand.

On a first pass, (18) appears to commit us to nothing more than times and people. But, if we try to symbolize it in first-order logic by, say,

(19) (t) $(Vsct \rightarrow (\exists t')(Bt't \& Vsmt'))$

problems begin to emerge. Let us suppose that Catherine is Saul's lover and Matthew is his boring cousin. Let us further suppose that Saul's mother has insisted that any time he visits Catherine he must first visit Matthew. Finally, interpret (18) as a report that Saul obeyed his mother. If (19) were an acceptable paraphrase of (18), Saul could have obeyed his mother by visiting Matthew just once before paying any visits to Catherine. I do not know about Saul's mother, but my mother would not put up with that. We cannot obtain an easy paraphrase of (18) by enlisting Boolos's "For every A, there is a B" either (TB, 431). For even this stronger sentence will not do:

(20) $(\exists f)[f \text{ is } 1-1 \& (t)(Vsct \rightarrow (Vsmf(t) \& Bf(x)t)]$

(There is a 1-1 correlation between the times of Saul's visits to Catherine and those of his visits to Matthew which mates every one of the former with an earlier one of the latter.) Saul could satisfy this by

worry that we cannot quantify over all sets, since doing so presupposes a universal set to serve as the domain of quantification. See my "Note on Interpreting Theories," Noûs, VIII, 3 (September 1974): 989 994 CC-0. In Public Domain. Gurukul Kangri Collection, Handwar

first visiting Matthew, then visiting Catherine and then visiting Matthew again (during the same trip) and thereafter visiting Matthew after visiting Catherine. As far as I can tell, whether we analyze (18) in terms of times, visits, or events, we must introduce some means for appropriately linking Saul's visits to Matthew with his visits to Catherine. One way of doing this is to use contextually determined time intervals. (For example, if Saul lived in New York and Catherine and Matthew both lived in Princeton, then the intervals in question might consist of Saul's entire visit to Princeton.) If we symbolize "i is an appropriate interval" by 'Ai' and "t and t' are part of i" by 'Ptt'i', then (21) seems to be a suitable explication of (18):

(21)
$$(t)(Vsct \rightarrow (\exists i)(\exists t')(Ai \& Ptt'i \& Bt't \& Vsmt'))$$

(For any time at which Saul visits Catherine, there is an appropriate interval containing a prior time at which he visits Matthew.) Given that Saul did visit Catherine, (21) not only commits us to times and people, it also commits us to time intervals.

Of course, someone else may prefer a treatment that eliminates the use of intervals. For example, if we introduce the relation "appropriately before," symbolized by 'A', then we can render (18) as

(22)
$$(t)(Vsct \rightarrow (\exists t')(At't \& Vsmt'))$$

and apparently avoid the commitment to intervals. But then we are left with an adverbial problem, i.e., that of explaining how (22) implies that Saul's visits to Matthew occurred before his visits to Catherine.

I do not mean to settle the example here, but I do wish to draw some morals from it. First, it should remind us that communication generally presupposes shared background beliefs and theories (such as beliefs concerning which visits to Matthew are appropriately prior to visits to Catherine). Hence, by virtue of their place within such systems of belief, our utterances may carry additional ontic commitments with them, which we frequently do not suspect until we subject them to logical analysis. Thus, the mere intelligibility of plural quantification and its prima facie lack of commitment to collections fails to be decisive.

Furthermore, as the competing analyses of both Geach-Kaplan and (18) illustrate, formalization is not simply a matter of considering candidate symbolizations in the light of our linguistic intuitions, but rather more a matter of applying broad linguistic, logical, and metaphysical theories. When there are competing candidate symbolizations, deciding between them may involve choosing between pervasive theories having different advantages and disadvantages on a variety of dimensions. Meaningful comparisons of competing pro-

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennal and eGangottin between posals might prove elusive. In any case, a company on between Boolos's proposal and the standard view of plural quantification is in order, although we have reason to doubt that this will produce decisive results.

Boolos's treatment has the obvious advantage of offering us a theory of ontic commitment which forestalls some commitments to classes.9 This leads to many benefits, including those I canvassed in my opening section. Of course, the standard treatment via classes

fails miserably on this comparison.

On the other hand, we should not forget that Boolos's approach has significant disadvantages too. It entails embracing second-order logic (over and above any set or class theory one might accept). If we accept Boolos's package, then one of the main objections to secondorder logic-its ontological commitment-would dissolve. But the other-its lack of proof procedure-would remain. I will not rehearse the reasons why this is a highly undesirable feature of second-order logic, since this has been very well done in a recent paper by Steven Wagner. 10

There are also some technical drawbacks to Boolos's proposal. So far at least, no one has been able to find a direct translation of second-order universal quantification in terms of plural quantification. [Even Boolos's rule for the existential case is not particularly natural. The plural "There are some things F" implies that there is at least one F, but the second-order existential quantifier has no first-order existential import. Thus, Boolos's rule must render " $(\exists F)Fa$ " as "Either $a \neq a$ or there are some things such that a is one of them." The translation in the other direction is simpler: "There are some things F. ... becomes $(\exists F)((\exists x)Fx \& ...)$.] Also, Boolos's procedure works only for monadic second-order quantification. To be sure, we can reduce the other cases to the monadic case by enlisting ordered n-tuples. Thus, we can construe an assertion of the form " $(\exists F)(\exists x)(\exists y)Fxy$ " as "There are some things such that $\langle x,y\rangle$ is one of them." But that means that the ontology of the theory in which the assertion is made must be closed under ordered pairing. This is no problem for set theory or even number theory, but it boosts the cost of applying Boolos's procedure to theories lacking devices for forming n-tuples. In sum, under Boolos's treatment there

10 See his "The Rationalist Conception of Logic," forthcoming in the Notre Dame Journal for Formal Logic. This paper also contains a valuable discussion of the issue of changing logic in order to achieve ontological economy.

⁹ It is important to realize that Boolos's criterion of ontic commitment differs from Quine's, since the latter is embedded in a logical theory that holds that ontic assessments make sense only with respect to referential first-order

are complications and ungainly asymmetries, which the treatment of

plural quantification by means of collections avoids.

Let me return to the topic with which we began this section: Boolos's circle. Despite Boolos's careful analyses, plural quantification remains problematic. We know something about its logic, but the question of its ontic commitment remains open. By taking it as a primitive for his logical theory, Boolos has put himself on the poor methodological footing of assuming that this issue is already closed.

The problematic nature of plural quantification is accentuated by Boolos's claims (discussed in section II) concerning synonymies. For, if he is correct, the plural quantifier is significantly equivocal: some substitutions for the dots in "There are some things such that . . . them . . ." give rise to plural quantifications that are irreducibly so, whereas others just produce variants on singular quantifications. Given that the latter occur only if certain formulas are second-orderequivalent, there is no complete set of complete rules for establishing, much less determining, whether a given quantifier is genuinely plural.

What is more, if there is such an equivocacy, then Boolos is worse off than the truth-functional logician is when dealing with the equivocacy attached to 'or'. For the latter can always define his disjunctions in terms of conjunction and negation, whereas Boolos still lacks a definition of plural existentials in terms of unequivocal plural universals. Furthermore, suppose that he tried to deal with the equivocacy as we usually deal with that attached to 'or'. He could start by pointing out that there are two senses to "There are some things such that" and that in one sense it just means "There is at least one thing such that." But where would he go from there? Obviously not to: In the other sense it means "There is a collection of things such that." But that seems to put him in the ironic position of saying that in the other sense it simply means what it says.

MICHAEL D. RESNIK

University of North Carolina/Chapel Hill

THE JOURNAL OF PHILOSOPHY

Digitized by Arya Samaj Foundation Chennai and eGangotri COMMENTS AND CRITICISM

BURGE'S NEW THOUGHT EXPERIMENT: BACK TO THE DRAWING ROOM

HEREAS Hilary Putnam has contended that "meanings ain't in the head," at least not entirely, Tyler Burge has gone even further, making the same claim about mental contents. In previous papers,1 Burge offered several arguments, based on twin-earth thought experiments, against individualistic individuation of types of mental states. Now he has devised a new argument.2 Assuming that kinds of states "are individuated by reference to intentional notions," he has given a new reason for claiming that "there are certain relations between an individual and the environment that are necessary to [his] having certain intentional notions" (709). This time Burge relies not on adaptations to intentional contents of Putnam's twin-earth fantasies concerning meaning, but on an original thought experiment. It features "not incomplete [linguistic] understanding or ignorance of specialized knowledge, but nonstandard theory" on the part of the relevant protagonists (709). Having elsewhere countered Burge's earlier arguments,3 here I will address only the new one. I will suggest that it depends, as do the intuitions that motivate it, on reading too much into the language we use to attribute propositional attitudes. In particular, I will challenge Burge's assumption that when we use a term in the 'that'-clause of an attitude attribution literally and correctly, we must be ascribing the notion expressed by the term to the content of the attitude.

¹ "Individualism and the Mental," *Midwest Studies in Philosophy*, IV (1979): 73–121, and "Other Bodies," in Andrew Woodfield, ed., *Thought and Object* (New York: Oxford, 1982).

² "Intellectual Norms and Foundations of Mind," this JOURNAL, LXXXIII, 12 (December 1986): 697–720. Unless otherwise indicated, all references to Burge are to this article.

Burge's earlier arguments depend on questionable intuitions (see fn 16 below) or on questionable descriptions of his examples. For instance, in "Individualism and the Mental," op. cit., Burge describes his well-known "tharthritis" example as involving "incomplete understanding of concepts" or "conceptual errors," even though it is more aptly described as involving incomplete understanding of words or linguistic errors. Indeed, sometimes he seems to equate the two, as when he says that "the error is linguistic or conceptual" (ibid., p. 82), though surely these are distinct. Burge offers no positive account of what it is to think with a concept that one incompletely understands. As for me, it is clear how one can use a word one incompletely understands, but I have no idea what it is to think with a concept that one incompletely understands, for I have no idea what it is to understand a concept over and above possessing it.

The new thought experiment is presented against the background of an original challenge which Burge poses to the dogma that necessary truths cannot be doubted. The truths he has in mind are those which "intuitively give the meaning of an empirically applicable term," such as

(1) Sofas are pieces of furniture [of such and such construction] made or meant for sitting.

He also challenges the related dogma that understanding a statement like (1) entails accepting it. These dogmas are supported by the following reasoning, which Burge tries to undermine: 'sofa' means piece of furniture . . . made for sitting; if one understands both phrases, to doubt (1) would be to doubt that sofas are sofas; therefore, one cannot doubt (1) while understanding it—recognizing its meaning entails recognizing its truth. Moreover, so the reasoning goes, "anyone who appears to be doubting that sofas are pieces of furniture . . . made for sitting cannot really be doing so" (701). Presumably the person is doubting something else instead, though Burge does not make this corollary of the reasoning explicit.⁴

To undermine this reasoning, Burge presents a provocative conception of individual and collective linguistic activity, according to which "synonymies are grounded in practice: the most competent speakers would use the two relevant expressions interchangeably" (701). He describes "a vast, ragged network of interdependence, established by patterns of deference which lead back to people who would elicit the assent of others" (702). Within this network occurs the dialectical process of giving explications of meaning, or "normative characterizations," which seeks an equilibrium between formulations and "archetypal applications." Burge insists that this dialectic concerns both "how correctly to characterize the relevant entities . . . and how to state the meaning of the [relevant] term; . . . examples are ineliminable from our procedures of meaning-giving" (705).⁵

5 "These disputes usually concern . . . [both] how correctly to characterize the relevant entities; whether all chairs have legs or must have legs; . . . [and] how to state the meaning of the term as such; whether according to (by) definition chairs have legs. . . . The second is typically pursued by trying to answer the first" (704/5). This process goes both ways, however: settling the question whether by

⁴ At one point Burge says that "the key idea is that some necessarily true thoughts or statements can be doubted" (698; my italics), as if there were no relevant difference between doubting thoughts and doubting statements. The difference, however, is relevant and important; for one can fully understand a thought and yet, because of linguistic ignorance, not fully understand a statement that expresses it. When Burge speaks of "what is doubted," he often fails to make clear whether he means a thought or a statement expressing that thought.

Burge's conception of this dialectic has two features essential to his critique of the reasoning outlined above. First, people can have object-level thoughts, e.g., thoughts about sofas, without possessing a correct normative characterization of the term in question, in this case 'sofa'. Second, although such characterizations, as arrived at through this dialectic, provide norms for understanding, their truth does not rest simply on agreement. People can agree and yet be wrong—general acceptance of a normative characterization may be premature. Even though the meaning of a term is constituted by social practice (it is dependent on usage and on people's critical reflection thereon), it need not be in anyone's head. Yet one's participation in this practice enables one to operate with the notion expressed by the term. In particular, someone with certain strange ideas about sofas can doubt a necessary truth like (1), even though he possesses the notion of sofa.

Such a person, A, is one of the protagonists in Burge's new thought experiment. It is designed to show, given the dubitability of meaning-giving normative characterizations, that "social practices are not the only or ultimate nonindividualistic factor in individuating mental states and events" (707). A, who has a normal mastery of English, understands the word 'sofa' and has acquired the normal truisms about sofas. However, he comes to doubt these truisms (e.g., he thinks that sofas would collapse under people's weight) and hypothesizes, contrary to (1), that sofas are really works of art or religious artifacts. He even devises an elaborate account of people's misconceptions about sofas. He is prepared to put his hypothesis to the test empirically, proposes some reasonable tests, and, ultimately, once their results come in, even acknowledges that he was mistaken.⁷

definition chairs have legs may lead not only to the conclusion that 'chair' implies having legs, but even to the conclusion that some of the things we have taken to be chairs, have been calling "chairs," are not chairs after all. As we will see, something analogous to this once happened in the case of 'witch', except that *no* so-called "witch" turned out to be one.

⁶ This phenomenon illustrates what R. D. Laing, in *The Politics of Experience* (New York: Ballantine Books: 1967), aptly describes as "conformity to a presence that is everywhere elsewhere" (p. 84), though his examples of such conformity are

⁷ Burge gives plenty of detail to make this story vivid and plausible. If you are not convinced that there could be widespread error about the function of common ing cats and feline robots. Burge remarks that the thought experiment can be made be adapted (709). Our criticism of the experiment will not exploit the choice of example.

Then there is *B*, *A*'s physically identical counterpart in another world, a world in which the things called "sofas" really are works of art or religious artifacts. Burge calls them "safos," but people in *B*'s world use 'sofa' to refer to them. Now it so happens that *B* hears counterparts of the truisms heard by *A*, but they are, unbeknownst to *B*, jokes, lies, or figurative utterances. *B* takes them literally and accepts them at first, taking safos to be pieces of furniture . . . made for sitting, but eventually develops doubts much like *A*'s. Unlike *A*, however, *B mistakenly* thinks that he is challenging common opinion—like everyone else, he correctly thinks that safos are works of art or religious artifacts.

According to Burge, the experiment shows that, even though A and B are physically similar and in phenomenologically similar situations, their 'sofa'-thoughts do not have the same contents. Whereas "A mistakenly thinks that sofas do not function primarily to be sat upon, B's counterpart thoughts do not involve the notion of sofa and could not correctly be ascribed with 'sofa' in oblique position"; A's thoughts can be so ascribed. In short, "A and B have different thoughts" (708).

Do they? Clearly A and B have thoughts about different things, sofas and safos respectively, but that is not at issue. Nor is Burge claiming that there is a difference between what A and B think about these things (sofas or safos, as the case may be), i.e., between the predicative contents of these thoughts. Rather, the difference is that, whereas "A has numerous mental events involving the notion of sofa, B's skepticism does not involve thinking of anything as a sofa" (708). So Burge is claiming that A thinks of the objects of the thoughts expressible or ascribable with the word form 'sofa' as sofas, whereas B thinks of the counterpart objects (his thoughts are expressible or ascribable in B's world with the word form 'sofa') as safos. We would attribute 'sofa'-thoughts to A with ascriptions of the form,

(2) A believes that sofas are. . . .

and we could, if we adopted the word 'safo', attribute 'sofa'-thoughts (i.e., thoughts expressible by the word form 'sofa' as used in B's world) to B with ascriptions of the form

(3) B believes that safos are. . .

Another way to put Burge's point, though he does not put it this way, would be to say that people in B's world use the form of (2) to

⁸ Burge uses the phrase 'word form' to avoid suggesting that the same word 'sofa' occurs in the languages of both worlds.

attribute beliefs to B, but would mean safo by 'sofa', thereby imput-

ing the notion of safo to B.

Underlying Burge's interpretation of this thought experiment is the assumption that, if one literally and correctly uses a term in the 'that'-clause of an attitude attribution, one is imputing to the subject the notion expressed by the term and is, further, including it in the content of the attitude being attributed.9 We commonly do this all right, but, I claim, not always. When using a form of ascription such as (2), it is our standard practice to use a word like 'sofas' both to refer to sofas, thereby expressing our notion of sofa, and to impute the same notion to the subject in his thoughts about sofas. Our standard practice must be modified, however, when the situation is not standard, as in Burge's thought experiment. For, in that situation, although A possesses the notion of sofa, the thought he expresses when he denies (1) by uttering (4),

(4) Sofas are not pieces of furniture . . . made for sitting.

does not involve the notion of sofa. To be sure, he possesses the notion of sofa and that is what the word 'sofa' means, even to him, but in using (4) to say something about sofas, he is not using the notion of sofa to pick out sofas. We may be inclined to use the form (2) to attribute beliefs to A, but this inclination can be explained by the fact that we are employing the notion of sofa to think of sofas when we use the word 'sofas' to refer to sofas. It is perfectly consistent for us to cancel the implication that A thinks of sofas in the way that we do. We can say, "A believes that sofas are . . . , not that he takes them to be sofas."

What makes plausible Burge's contention that ascriptions of the form of (2) literally describe A's beliefs (including the belief that sofas are not pieces of furniture . . . made for sitting) and impute to him the notion of sofa is that A does possess the notion of sofa, he does associate this notion with the word 'sofa', and he does use that word to refer to sofas. However, Burge does not consider the possibility that, when A uses 'sofa' to refer to sofas, he does not intend to be using the word literally. For, although 'sofa' means piece of furniture . . . made for sitting, even to him, he does not take the things to which he is using 'sofa' to refer to be sofas. If he utters (4), he means that so-called "sofas" (what everybody else calls "sofas") are not pieces of furniture. . . made for sitting. He is exploiting the

⁹ More broadly, Burge assumes that our use of 'that'-clauses in attitude attributions inherently embodies a (correct) principle of psychological state individuation. I challenge this assumption in Thought and Reference, op. cit., ch. 10, and in "De re Belief and Methodological Solipsism," in Woodfield, ed., op. cit.

common, but in his view mistaken, belief that those items are sofas. Indeed, he might well go on to say that he does not take those items to be sofas. Only then would he be using 'sofa' literally. 10

This way of describing the situation with A is perfectly natural, as can be seen if we change the example from sofas to witches. Imagine being in New England in the period when deranged women were called witches, where 'witch' meant woman in league with the Devil (or something of the sort). Suppose you began to doubt common opinion and came to believe that these women are not in league with the Devil. Your fellow New Englanders would use (5):

(5) S believes that witches are not in league with the Devil.

to attribute this belief to you, but we would not—unless we put 'witches' in scare quotes. Your fellow New Englanders take it to be a necessary truth that witches are in league with the Devil, which they take you to reject, but we do not suppose there are any witches at all. Like you, we reject their belief in witches, though not their notion of witch. Yet Burge's position seems to require that (5) attribute a belief literally and correctly and implies that your belief involves the notion of witch. But then what you believe is the negation of a necessary truth, 11 in which case it could not be true. Clearly, a better course is to say that (5) attributes to you a belief about so-called "witches," a belief which does not involve the notion of witch, and that you disagree with your fellow New Englanders about what these individuals are. This is not to deny that you possess the notion of witch; indeed, it figures in your belief that there are no witches.

A parallel situation would obtain if A were right about sofas, a possibility Burge does not disallow.¹² Then, if people were to de-

Burge might contend that it is not a necessary truth but false that witches are in league with the Devil, and that witches are (really) insane women, but then he would be implying that witches exist.

¹² Burge's conception of linguistic practice explicitly allows for the possibility that we could be systematically mistaken about a certain class of objects, so that the items

on the part of the audience [see Kent Bach and Robert M. Harnish, Linguistic Communication and Speech Acts (Cambridge, MA: MIT Press, 1979), pp. 85–87]. The situation described here is somewhat analogous to those examples of referential use of definite descriptions where the speaker does not believe the referent to satisfy the description he is using but is exploiting the fact that his audience believes that it does. Suppose you use 'the man drinking a martini' to refer to someone drinking Sprite in a martini glass. You are not using the phrase literally, although you are exploiting its literal meaning and the fact that your audience would take the man to be drinking a martini. Since you intend to be taken literally, however, the analogy with A's use of 'sofa' is imperfect—he does not intend to be taken literally. See my Thought and Reference, op. cit., ch. 6, for further discussion of the referential case and of the referential/attributive distinction.

scribe A as thinking that sofas are not pieces of furniture. . . made for sitting, but works of art or religious artifacts, they would be attributing to him a belief that conflicts with the common conception of sofas, but in this case the common conception would be a mistaken conception of the items called "sofas." Burge needs to show that there is a relevant difference between the witch case, where so-called "witches" are not really witches, and his sofa case, where so-called "sofas" really are sofas. Yet it seems that the only relevant difference concerns the existence of the objects in question (i.e., whether so-called Ks really are Ks), and the contents of a person's beliefs and the notions involved in them should not depend on the existence or nonexistence of the objects in question. What applies to the skeptic about witches should apply equally to the skeptic about sofas. 13

Before elaborating my rejoinder to Burge's interpretation of the thought experiment, I should mention that Burge does anticipate certain other rejoinders. They all challenge his claim that A literally doubts what he appears to doubt, and suggest that A does not really possess the notion of sofa but does possess some other notion instead. One of these rejoinders¹⁴ is similar to mine, except that it refuses to impute to A the notion of sofa: "A thinks only that what most people think of as sofas are works of art or religious artifacts" (711). Burge replies, "We may assume that A would say that what most people think of as sofas are sofas. . . . What he questions, quite explicitly we may imagine, is not whether these things—or what people think of as sofas—are sofas, but whether sofas are what

we call 'Ks' are not Ks at all, as we would acknowledge if we knew their true nature. In "Other Bodies," op. cit., Burge allows for such a possibility in an example involving water, though he speaks of it as "epistemic" (p. 114). With sofas, however, it is also metaphysically possible that they not be pieces of furniture . . . made for sitting.

A, whatever he may say, "has no thoughts literally attributable to him by using 'sofa' in oblique position, but only metalinguistic thoughts about the word 'sofa' " (710), or that he has thoughts involving only a "reduced" notion of sofa (711), e.g., a phenomenal one.

¹⁸ In this regard consider Burge's contention that, if a doubt like A's "were to prove well founded, the conventional meaning of 'sofa' would be forced to change. But, despite the change, it might remain appropriate, before and after the change, to attribute propositional attitudes involving the notion of sofa. Both before and over whether all and only sofas are furnishings of a certain structure made or meant Moreover, the meaning of the word 'witch' did not change when people became enlightened about so-called "witches."

people think they are" (711; my italics). Of course, what Burge claims that we may assume or imagine is just what is at issue. Instead of arguing for it, Burge baldly asserts,

There is simply nothing in our ordinary practices that precludes our taking A as literally entertaining the doubts I have ascribed. Whatever impetus there is behind the objections derives not from antecedent practice, but from the feeling that there must be some way of resisting the ascriptions. This feeling derives from habits that stem from background philosophical doctrine (712).

For my part, this feeling derives (even if I am a diehard individualist) from resisting the habits that stem from a background philosophical doctrine taken for granted by Burge. He assumes that, for an ascription of a thought to be literal, each term occurring in the 'that'-clause must be used to express an element in the content of the thought being ascribed. But that is simply false. The reason, as noted earlier, is that the ascriber can use such a term to express how he, the ascriber, is thinking of the object of the subject's thought. Thus, one can say of A,

(6) A believes that sofas are . . religious artifacts.

and be speaking literally and correctly, even though one is using 'sofas' to refer to sofas and not to express how A thinks of the objects of the thought being ascribed. It is being denied not that A possesses the notion of sofas but only that this notion figures in the content of the thought in question. However, if one said of A,

(7) A believes that sofas do not exist.

one would be using 'sofas' to impute the notion of sofa to A, but one would not be referring to sofas, since one is not attributing to A the belief that they, the things everyone else takes to be sofas, do not exist. Notice also that, if A uttered (8),

(8) [I believe that] sofas are . . . religious artifacts.

he would not be speaking literally, since even though he is using 'sofas' to refer to what in fact are sofas, he does not take them to be sofas and is thinking of them as what most people think of as sofas (or as so-called "sofas"). However, if A uttered (9),

(9) [I believe that] sofas do not exist.

¹⁵ I develop this and related points about belief sentences in "De re Belief and Methodological Solipsism," op. cit., pp. 129–131, and in Thought and Reference, op. cit., ch. 10.

Digitized by Arya Samaj Foundation Chennai and eGangotri he would be speaking literally. He would not be denying the existence of the things that he takes to be . . . religious artifacts.

Finally, recall that part of Burge's purpose is to show that one can doubt necessary truths that one understands. Thus, he insists that

(10):

(10) A believes that sofas are not pieces of furniture . . . made for sitting.

is literally true. I would grant that, but only in the way that (11):

(11) A believes that sofas are not sofas.

is literally true, namely, where the first occurrence of 'sofas' is used to refer to sofas and only the second is used to express part of the content of A's thought. And (11) is literally true in this way, as is evident if one appended 'but . . . religious artifacts' to (11).

Let us conclude our discussion of the thought experiment by considering the situation of A's counterpart B, whom we have been neglecting. Recall that B is relevantly similar to A. In particular, his experience of sofa-like items and his exposure to uses of the word form 'sofa' are the same. But, of course, in B's world these objects are safos and this word form means safo. B is right about the objects but wrong about popular opinion. Moreover, according to Burge, B possesses the notion of safo, not that of sofa, and B's thoughts "do not involve the notion of sofa and could not correctly be ascribed with 'sofa' in oblique position"; A's thoughts can be so ascribed. A and B have different thoughts, and thus the thought experiment refutes individualism.

Burge's only reason for denying that B possesses the notion of sofa seems to be, however, that in B's world the word form 'sofa' means safo, as in ascriptions of the form

(12) B believes that sofas are. . .

Now, of course, people in B's world do use 'sofas' to refer to safos, even in the context of (12), and use it to express the notion of safo. But is does not follow that they would use it, as in (12), to impute the notion of safo to B. Assuming they understood his views, they would realize that to him 'sofa' means not safo but piece of furniture . . . made for sitting. And, although this is what it means to him, if he uttered (13),

(13) Sofas are not pieces of furniture . . . made for sitting.

he would be using 'sofa' to express not this notion but rather socalled "sofa" (as we saw earlier, the same goes for A). Moreover,

Digitized by Arya Samaj Foundation Chennai and eGangotri even though B uses 'sofas' to refer to safos, just as everyone else does, and even though he thinks that they are . . . religious artifacts, he does not intend to be using 'sofas' literally (as it is used in his world). For he thinks that 'sofa' means piece of furniture . . . made for sitting, and he does not take safos (which everyone in his world uses 'sofa' to refer to) to be pieces of furniture . . . made for sitting. It so happens that he is unwittingly using 'sofa' to refer to what are literally sofas in the sense of 'sofa' operative in his world (safos), but that is accidental; he does not intend to be using 'sofa' in that sense.

On close examination, Burge's new thought experiment loses its momentous import. That is why we did not consider variants of it, although we could have. For, as Peter Unger¹⁶ has shown, tinkering with the details of a thought experiment can dramatically change one's intuitions. We could have asked, for example: What if roughly half the items called "sofas" are pieces of furniture . . . made for sitting and the rest are works of art or religious artifacts? or What if before 1900 the items called "sofas" were works of art or religious artifacts and after 2000 this will again be the case? We did not take up such questions because, rather than vary the details, as Unger might have done, we considered it sufficient to inspect them as given. The intuitions that support Burge's radical position may be initially plausible, but they become undermined once we look carefully at how the relevant terms are used in various contexts. Then we find that nothing radical can be inferred about propositional attitudes from Burge's new experiment with the language used to attribute them.

KENT BACH

San Francisco State University

¹⁶ See Peter Unger, *Philosophical Relativity* (Minneapolis: University of Minnesota Press, 1984), ch. V. Unger observes that, even in the familiar versions of these experiments, there is a tension between "existence beliefs" and "property beliefs." For example, although we all believe that there are cats and that cats are animals, what if the things we have been calling "cats" are actually robots controlled from Mars? Either there are no cats or cats are not animals. My intuitions favor the former option, Putnam's the latter. Unger plays no favorites, thinking that a good case can be made for either option.

AARNIO, AULIS: The Rational as Reasonable: A Treatise on Legal Justification.

Dordrecht: Nijhoff, 1986. xix, 276 p. \$59.00.

ADAMS, D. J.: Diderot, Dialogue and Debate. Wolfeboro, NH: Longwood, 1986. 216 p. \$25.00.

ADAMS, MARILYN MCCORD: William Ockham, Vol. I & II. Notre Dame, IN: University Press, 1987. xx, 1402 p. \$90.00.

ADAMS, ROBERT M.: The Virtue of Faith and Other Essays in Philosophical Theology. New York: Oxford, 1987. 270 p. Cloth \$29.95, paper \$12.95. ALEXANDER, RICHARD D.: The Biology of Moral Systems. Hawthorne, NY:

Aldine, 1987. xx, 307 p. Cloth \$34.95, paper \$16.95.

ALEXANDER, THOMAS M.: John Dewey's Theory of Art, Experience & Nature: The Horizons of Feeling. Albany: SUNY Press, 1987. xxiv, 325 p. Cloth \$44.50, paper \$16.95.

ARBIB, MICHAEL A., and MARY B. HESSE: The Construction of Reality. New York: Cambridge, 1986. xii, 286 p. \$29.95.

ARONSON, RONALD: Sartre's Second Critique. Chicago: University Press, 1987. xvii, 255 p. Cloth \$32.50, paper \$13.95.

ARTHUR, C. J.: Dialectics of Labour: Marx and his Relation to Hegel. New York: Basil Blackwell, 1987. 182 p. \$39.95.

BAIGENT, MICHAEL, RICHARD LEIGH, and HENRY LINCOLN: The Messianic Legacy. New York: Holt, 1987. xvi, 364 p. \$19.95.

BALL, TERENCE: Idioms of Inquiry: Critique and Renewal in Political Science. Albany: SUNY Press, 1987. vi, 255 p. Cloth \$44.50, paper \$14.95.

BARRY, NORMAN P.: Classical Liberalism and Libertarianism. New York: St. Martin's, 1987. xi, 215 p. \$29.95.

BARTON, G. EDWARD, ROBERT C. BERWICK, and ERIC SVEN RISTAD: Computational Complexity and Natural Language. Cambridge, MA: MIT Press, 1987. xii, 335 p. \$24.95.

BARWISE, JON, and JOHN ETCHEMENDY: The Liar: An Essay in Truth and Circularity. New York: Oxford, 1987. xii, 185 p. \$19.95.

BEDAU, HUGO ADAM: Death is Different. Boston: Northeastern UP, 1987. xii, 307 p. \$25.00.

BEISER, FREDERICK C.: The Fate of Reason: German Philosophy from Kant to Fichte. Cambridge, MA: Harvard, 1987. xi, 395 p. \$30.00.

BENNETT, JANE: Unthinking Faith and Enlightenment: Nature and the State in a Post-Hegelian Era. New York: NYU Press, 1987. viii, 166 p. \$30.00.

BERRY, CHRISTOPHER J.: *Human Nature*. Atlantic Highlands, NJ: Humanities, 1986. xiv, 162 p. Cloth \$29.95, paper \$8.50.

BIRD, GRAHAM: William James. New York: Methuen, 1987. ix, 221 p. \$39.95. BLACKBURN, SIMON: Spreading the Word. New York: Oxford, 1986. xi, 365 p. \$12.50.

BLAZER, WOLFGANG, C. ULISES MOULINS and JOSEPH D. SNEED: An Architectonic for Science: The Structuralist Program. Dordrecht: Nijhoff, 1987. xxxvii, 431 p. \$99.00.

- Digitized by Arya Samaj Foundation Chennai and eGangotri BOGDAN, RADU J., ed.: *Jaakko Hintikka*. Dordrecht: Reidel, 1987. ix, 328 p. \$89.00.
- BORDO, SUSAN R.: The Flight to Objectivity: Essays on Cartesianism and Culture. Albany: SUNY Press, 1987. xi, 145 p. Cloth \$34.50, paper \$12.95.
- BOWERS, C. A.: Elements of a Post-Liberal Theory of Education. New York: Teachers College, 1987. ix, 187 p. Cloth \$27.95, paper \$16.95.
- BOYLE, JOAN M., and JAMES E. MORRISS: The Mirror of Time: Images of Aging and Dying. Westport, CT: Greenwood, 1987. xiii, 225 p. \$35.00.
- BRATMAN, MICHAEL E.: Intention, Plans, and Practical Reason. Cambridge, MA: Harvard, 1987. viii, 200 p. \$25.00.
- BRAYBROOKE, DAVID: Meeting Minds. Princeton, NJ: University Press, 1987. xi, 344 p. Cloth \$35.00, paper \$12.50.
- BRIEN, KEVIN M.: Marx, Reason, and the Art of Freedom. Philadelphia: Temple UP, 1987. xvi, 260 p. \$32.95.
- BROADIE, ALEXANDER: Introduction to Medieval Logic. New York: Oxford, 1987. 150 p. \$38.00.
- BROWN, HANBURY: The Wisdom of Science: Its Relevance to Culture and Religion. New York: Cambridge, 1986. viii, 194 p. Cloth \$32.50, paper \$13.95.
- BROWN, HAROLD I.: Observation and Objectivity. New York: Oxford, 1987. xii, 255 p. \$24.95.
- BRUNDELL, BARRY: Pierre Gassendi: From Aristotelianism to a New Natural Philosophy. Dordrecht: Reidel, 1987. x, 251 p. \$64.00.
- BUCKLEY, MICHEAL J.: At the Origins of Modern Atheism. New Haven, CT: Yale, 1987. viii, 445 p. \$27.50.
- BUCKNELL, RODERICK S., and MARTIN STUART FOX: The Twilight Language: Explorations in Buddhist Meditation and Symbolism. New York: St. Martin's, 1987. xii, 233 p. \$27.50.
- CALLAHAN, DANIEL: Setting Limits: Medical Goals in an Aging Society. New York: Simon and Schuster, 1987. 256 p. \$18.95.
- CALLICOT, J. BAIER, ed.: Companion to A Sand County Almanac: Interpretive and Critical Essays. Madison: Wisconsin UP, 1987. x, 308 p. \$22.50.
- CANNING, JOSEPH: The Political Thought of Baldus de Ubaldis. New York: Cambridge, 1987. xi, 300 p. \$49.50.
- CARE, NORMAN S.: On Sharing Fate. Philadelphia: Temple UP, 1987. ix, 245 p. \$29.95.
- CARLSNAES, WALTER: Ideology and Foreign Policy: Problems of Comparative Conceptualization. New York: Basil Blackwell, 1987. xi, 234 p. \$39.95.
- CARR, DAVID: Interpreting Husserl. Dordrecht: Nijhoff, 1987. x, 303 p. \$70.00.
- CASEY, EDWARD S.: Remembering: A Phenomenological Study. Bloomington: Indiana UP, 1987. xiv, 362 p. Cloth \$47.50, paper \$18.50.
- CATALANO, JOSEPH S.: A Commentary on Jean-Paul Sartre's Critique of Dialectical Reason, Vol. 1, Theory of Practical Ensembles. Chicago: University Press, 1987. x, 282 p. Cloth \$40.00, paper \$15.95.
- CAVELL, STANLEY: Disowning Knowledge in Six Plays of Shakespeare. New York: Cambridge, 1987. x, 226 p. Cloth \$34.50, paper \$10.95.

Digitized by Arya Samai Foundation Chennal and Chamber From Aristotle to CHAMBLISS, J. J.: Educational Theory as Theory of Condamon From Aristotle to Dewey. Albany: SUNY Press, 1987. viii, 172 p. Cloth \$39.50, paper \$14.95.

CHAN, WING-TSIT: Chu Hsi: Life and Thought. New York: St. Martin's, 1987.

xii, 212 p. \$32.50.

CHARGAFF, ERWIN: Serious Questions: An ABC of Skeptical Reflections. Cambridge, MA: Birkhäuser, 1987. viii, 261 p. \$40.00. CHILD, JAMES W.: Nuclear War: The Moral Dimension. New Brunswick, NJ:

Transaction Books, 1987. 197 p. Cloth \$16.95, paper \$8.95.

CHISHOLM, R. M.: Brentano and Intrinsic Value. New York: Cambridge, 1986. xiii, 105 p. Cloth \$24.95, paper \$10.95.

CHRISTIAN, SR., WILLIAM A.: Doctrines of Religious Communities: A Philosophical Study. New Haven CT: Yale, 1987. vii, 234 p. \$25.00.

CODE, LORRAINE: Epistemic Responsibility. Hanover, NH: UP of New England, 1987. xi, 272 p. \$28.00

COHEN, BERNARD LANDE: Economics Without Ideology. New York: Philosophical Library, 1987. viii, 213 p. \$20.00.

COLEMAN, JAMES S.: Individual Interests and Collective Action. New York: Cambridge, 1986. ix, 384 p. \$54.50.

COLLINS, ARTHUR W.: The Nature of Mental Things. Notre Dame, IN: University Press, 1987. xviii, 180 p. \$22.95.

COLLINSON, DIANE: Fifty Major Philosophers: A Reference Guide. New York: Methuen, 1987. 170 p. Cloth \$29.95, paper \$11.95.

COXON, A. H.: The Fragments of Parmenides. Wolfeboro, NH: Van Gorcum, 1986. vii, 277 p. \$30.00.

CRAIG, EDWARD: The Mind of God and the Works of Man. New York: Oxford, 1987. x, 353 p. \$49.00.

CRONK, GEORGE: The Philosophical Anthropology of George Herbert Mead. New York: Lang, 1987. ix, 153 p. \$26.00.

CUTSINGER, JAMES S.: The Form of Transformed Vision: Coleridge and the Knowledge of God. Macon, GA: Mercer UP, 1987. xix, 136 p. \$24.95.

DAMROSCH, DAVID: The Narrative Covenant: Transformations of Genre in the Growth of Biblical Literature. San Francisco: Harper & Row, 1987. x, 352 p. \$21.95.

DASCAL, MARCELO: Leibniz: Language, Signs and Thought. Amsterdam: Benjamins, 1987. \$36.00.

DAVIDSON, HERBERT A.: Proofs for Eternity, Creation and the Existence of God in Medieval Islamic and Jewish Philosophy. New York: Oxford, 1987. xii, 428 p. \$37.50.

DAVIE, GEORGE: The Crisis of the Democratic Intellect. Totowa, NJ: Barnes & Noble, 1987. 283 p. \$28.50.

DENNETT, DANIEL C.: The Intentional Stance. Cambridge, MA: MIT Press, 1987. xi, 388 p. \$25.00.

DESMOND, WILLIAM: Desire, Dialectic, and Otherness: An Essay on Origins. New Haven, CT: Yale, 1987. x, 259 p. \$26.50.

DE SOUSA, RONALD: The Rationality of Emotion. Cambridge, MA: MIT Press, 1987, xix, 373 p. \$25.00.

DICKEY, LAURENCE: Hegel: Religion, Economics, and the Politics of Spirit, 1770-1807. New York: Cambridge, 1987. xvi, 459 p. \$39.50.

DOOB, LEONARD W.: Slightly Beyond Skepticism: Social Science and the Search for Morality. New Haven, CT: Yale, 1987. ix, 319 p. \$25.00.

DOYAL, LEN, and ROGER HARRIS: Empiricism, Explanation, and Reality. New York: Methuen, 1987. xiii, 200 p. \$44.00.

EDIE, JAMES M.: Edmund Husserl's Phenomenology: A Critical Commentary. Bloomington: Indiana UP, 1987. xviii, 150 p. Cloth \$27.50, paper \$9.95.

EDIE, JAMES M.: William James and Phenomenology. Bloomington: Indiana UP, 1987. xiii, 111 p. Cloth \$25.00, paper \$9.95.

ELSTER, JON, ed.: The Multiple Self. New York: Cambridge, 1986. 269 p. \$29.95.

EMERSON, RALPH WALDO: The Collected Works of Ralph Waldo Emerson, Vol. IV: Representative Men. Cambridge, MA: Harvard, 1987. xcii, 409 p. \$30.00.

ESCHER, NICHOLAS: Forbidden Knowledge and Other Essays on the Philosophy of Cognition. Boston: Reidel, 1987. xiv, 154 p. \$54.00.

ESPOSITO, JOSEPH L.: The Obsolete Self: Philosophical Dimensions of Aging. Berkeley: California UP, 1987. vi, 245 p. \$25.00.

EWIN, R. E.: Liberty, Community, and Justice. Totowa, NJ: Rowman and Little-field, 1987. vii, 219 p. Cloth \$32.50, paper \$14.95.

FAIN, HASKELL: Normative Politics and the Community of Nations. Philadelphia: Temple UP, 1987. xii, 244 p. \$32.95.

FAY, BRIAN: Critical Social Science. Ithaca, NY: Cornell, 1987. x, 242 p. Cloth \$35.00, paper \$12.95.

FERREIRA, JAMIE M.: Skepticism and Reasonable Doubt: The British Naturalist Tradition in Wilkins, Hume, Reid, and Newman. New York: Oxford, 1987. xii, 255 p. \$39.95.

FINNIS, JOHN, JOSEPH M. BOYLE, JR., and GERMAIN GRISEZ: Nuclear Deterence, Morality and Realism. New York: Oxford, 1987. xv, 429 p. \$39.95.

FISHER, WALTER R.: Human Communication as Narration: Toward a Philosophy of Reason, Value, and Action. Columbia: South Carolina UP, 1987. xiv, 201 p. \$24.95.

FITCH, G. W.: Naming and Believing. Dordrecht: Reidel, 1987. x, 215 p. \$44.50

FLAGE, DANIEL, F.: Berkeley's Doctrine of Notions: A Reconstruction Based on His Theory of Meaning. New York: St. Martin's, 1987. 226 p. \$29.95.

FLATHMAN, RICHARD E.: The Philosophy and Politics of Freedom. Chicago: University Press, 1987. x, 360 p. Cloth \$42.50, paper \$16.95.

FODOR, JERRY A.: Psychosemantics: The Problem of Meaning in the Philosophy of Mind. Cambridge, MA: MIT Press, 1987. xiii, 171 p. \$19.95.

FREDE, MICHAEL: Essays in Ancient Philosophy. Minneapolis: Minnesota UP, 1987. xxvii, 382 p. Cloth \$32.50, paper \$13.95.

FRINGS, MANFRED S.: Philosophy of Prediction and Capitalism. Dordrecht: Nijhoff, 1987. 147 p. \$39.00.

FUCHS, JOSEF: Christian Morality: The Word Becomes Flesh. Washington: Georgetown UP, 1987. 212 p. Cloth \$19.95, paper \$11.95.

The Park Street

GANDHI, MAHATMA: The Moral and Political Writings of Mahatma Gandhi, RAGHAVAN IYER, ed. New York: Oxford, 1987. Vol. II, xxii, 678 p. \$89.00. Vol III, xx, 641 p. \$99.00.

GAY, PETER: A Godless Jew: Freud, Atheism, and the Making of Psychoanalysis. New Haven, CT: Yale, 1987. xvii, 812 p. \$17.95.

GELLNER, ERNEST: Culture, Identity, and Politics. New York: Cambridge, 1987. viii, 189 p. Cloth \$39.50, paper \$12.95.

GIBBINS, PETER: Particles and Paradoxes: The Limits of Quantum Logic. New York: Cambridge, 1987. xi, 181 p. Cloth \$34.50, paper \$11.95.

GLYMOUR, CLARK, et al.: Discovering Causal Structure: Artificial Intelligence, Philosophy of Science, and Statistical Modeling. Orlando, FL: Academic, 1987. xvii, 394 p. \$39.50.

GOCHET, PAUL: Ascent to Truth: A Critical Examination of Quine's Philosophy. Munich: Philosophia, 1987. 201 p. 59 DM.

GODFREY, JOSEPH J.: A Philosophy of Human Hope. Dordrecht: Nijhoff, 1987. xiii, 272 p. \$50.50.

GOLDKIND, STUART: Machines and Intelligence. Westport, CT: Greenwood, 1987. x, 138 p. \$29.59.

GORDON, ROBERT M.: The Structure of Emotions. New York: Cambridge, 1987. xiv, 161 p. \$29.95.

GREENAWALT, KENT: Religious Convictions and Political Choice. New York: Oxford, 1987. xii, 266 p. \$29.95.

GREENAWALT, KENT: Conflicts of Law and Morality. New York: Oxford, 1987. xii, 383 p. \$29.95.

GRIFFIN, JAMES: Well-Being: Its Meaning, Measurement and Moral Importance. New York: Oxford, 1987. xii, 412 p. \$48.00.

GRINNELL, FREDERICK: The Scientific Attitude. Boulder, CO: Westview, 1987. xvii, 141 p. Cloth \$29.95, paper \$13.50.

GRUNDLEHNER, PHILIP: The Poetry of Friedrich Nietzsche. New York: Oxford, 1987. xxv, 359 p. \$24.95.

GRUNEBAUM, JAMES O.: Private Ownership. New York: Methuen, 1987, ix, 213 p. \$35.00.

GUTMANN, AMY: Democratic Education. Princeton, NJ: University Press, 1987. xii, 321 p. \$19.95.

GUTTENPLAN, SAMUEL: The Languages of Logic. New York: Basil Blackwell, 1987. x, 339 p. Cloth \$34.95, paper \$16.95.

HAJIME, TAKEUCHI: Philosophy as Metanoetics. Berkeley: California UP, 1987. lxii, 316 p. \$42.00.

HAMBLIN, C. L.: Imperatives. New York: Basil Blackwell, 1987. ix, 262 p. \$45.00.

HAMLYN, D. W.: A History of Western Philosophy. New York: Viking Penguin, 1987. 345 p. \$19.95.

HAMPTON, JEAN: Hobbes and the Social Contract Tradition. New York: Cambridge, 1987. xii, 299 p. \$42.50.

HANKEY, W. J.: God in Himself: Aquinas' Doctrine of God as Expounded in the Summa Theologiae. New York: Oxford, 1987. viii, 196 p. \$42.00.

- in the Theater and Philosophy of Gabriel Marcel (1889-1973). Lanham, MD: UP of America, 1987. xxiv, 224 p. \$24.25.
- HARLAND, RICHARD: Superstructuralism. New York: Methuen, 1987. x, 213 p. Cloth \$37.50, paper \$10.95.
- HARRE, ROM: Varieties of Realism: A Rationale for the Natural Sciences. New York: Basil Blackwell, 1987. viii, 375 p. \$34.95.
- HARRIS, ERROL E.: Formal, Transcendental, and Dialectical Thinking: Logic and Reality. Albany: SUNY Press, 1987. xii, 289 p. Cloth \$34.50, paper \$12.95.
- HEIM, MICHAEL: Electric Language: A Philosophical Study of Word Processing. New Haven, CT: Yale, 1987. xi, 305 p. \$19.95.
- HELLER, AGNES: Beyond Justice. New York: Basil Blackwell, 1987. vi, 346 p. \$39.95.
- HIGGINS, KATHLEEN MARIE: Nietzsche's Zarathustra. Philadelphia: Temple UP, 1987. xxi, 306 p. \$34.95.
- HILMY, STEPHEN S.: The Later Wittgenstein: The Emergence of a New Philosophical Method. New York: Basil Blackwell, 1987. viii, 340 p. \$49.95.
- HIXSON, RICHARD F.: Privacy in a Public Society: Human Rights in Conflict. New York: Oxford, 1987. xvi, 255 p. \$19.95.
- HOFFMAN, PIOTR: Doubt, Time, Violence. Chicago: University Press, 1987. xix, 146 p. \$25.00.
- HOLBROOK, DAVID: Evolution and the Humanities. New York: St. Martin's, 1987. 228 p. \$32.50.
- HONORE, TONY: Making Law Bind. New York: Oxford, 1987. vii, 274 p. \$57.00.
- HOOKER, C. A.: A Realistic Theory of Science. Albany: SUNY Press, 1987. vii, 479 p. Cloth \$39.50, paper \$14.95.
- HORWICH, PAUL: Asymmetries in Time: Problems in the Philosophy of Science. Cambridge, MA: MIT Press, 1987. ix, 218 p. Cloth \$22.50, paper \$9.95.
- HOULGATE, STEPHEN: Hegel, Nietzsche and the Criticism of Metaphysics. New York: Cambridge, 1987. xviii, 300 p. \$39.50.
- HUSSERL, EDMUND: Aufsätze und Vorträge (1911–1921). THOMAS NENON and HANS RAINER SEPP, eds. Dordrecht: Nijhoff, 1987. xxxix, 408 p. \$134.00.
- HUSSERL, EDMUND: Vorlesungen über Bedeutungslehre Sommersemester 1908. URSULA PANZER, ed. Dordrecht: Nijhoff, 1987. xxix, 265 p. \$92.50.
- IMBODEN, ROBERTA: From the Cross to the Kingdom. San Francisco: Harper & Row, 1987. xvi, 134 p. \$17.95.
- INGRAM, DAVID: Habermas and the Dialectic of Reason. New Haven, CT: Yale, 1987. xvii, 263 p. \$24.50.
- JACKENDOFF, RAY: Consciousness and the Computational Mind. Cambridge, MA: MIT Press, 1987. xvi, 356 p. \$27.50.
- JAMES, WILLIAM: The Works of William James: Essays, Comments, and Reviews. FREDSON BOWERS, ed. Cambridge, MA: Harvard, 1987. xxxix, 785 p. \$60.00.
- JARDINE, N.: The Fortunes of Inquiry. New York: Oxford, 1987. 144 p. \$34.00.

JOHNSON, JAMES TURNER: The Quest for Peace: Three Moral Traditions in Western Cultural History. Princeton, NJ: University Press, 1987. xx, 300 p. \$30.00.

The same of the sa

JOHNSON, MARK: The Body in the Mind. Chicago: University Press, 1987. xxxviii, 233 p. \$27.50.

JOHNSTON, MARK D.: The Spiritual Logic of Ramon Llull. New York: Oxford, 1987. xii, 336 p. \$65.00.

JONES, RICHARD H.: Science and Mysticism. Cranbury, NJ: Associated University Presses, 1986. 266 p. \$20.00.

KALSO, MARIE-LUISE SCHUBERT: Meinong's Theory of Knowledge. Dordrecht: Nijhoff, 1987. xiv, 128 p. \$36.00.

KASHAP, PAUL S.: Spinoza and Moral Freedom. Albany: SUNY Press, 1987. xvi, 198 p. Cloth \$34.50, paper \$12.95.

KELLY, CHRISTOPHER: Rousseau's Exemplary Life: The Confessions as Political Philosophy. Ithaca, NY: Cornell, 1987. xiv, 262 p. \$29.95.

KINNEAVY, JAMES L.: Greek Rhetorical Origins of Christian Faith: An Inquiry. New York: Oxford, 1987. xi, 186 p. \$29.95.

KITTAY, EVA FEDER: Metaphor: Its Cognitive Force and Linguistic Structure. New York: Oxford, 1987. 358 p. \$59.00.

KLADIMAN, STEPHEN, and TOM L. BEAUCHAMP: *The Virtuous Journalist*. New York: Oxford, 1987. x, 246 p. \$19.95.

KLEMAN, MARK: A Guide to Critical Legal Studies. Cambridge, MA: Harvard, 1987. vi, 360 p. \$30.00.

KLUBACK, WILLIAM: Eric Weil: A Fresh Look at Philosophy. Lanham, MD: UP of America, 1987. viii, 186 p. \$23.00.

KOLB, DAVID: The Critique of Pure Modernity: Hegel, Heidegger, and After. Chicago: University Press, 1987. xvii, 316 p. \$25.00.

KOLENDA, KONSTANTIN: Cosmic Religion: An Autobiography of the Universe. Prospect Heights, IL: Waveland, 1987. xiii, 138 p. \$7.95.

KRAGH, HELGE: An Introduction to the Historiography of Science. New York: Cambridge, 1987. viii, 235 p. \$32.50.

KROIS, JOHN MICHAEL: Cassirer: Symbolic Forms and History. New Haven, CT: Yale, 1987. xvii, 262 p. \$25.00.

KROLICK, SANFORD: Recollective Resolve. Macon, GA: Mercer UP, 1987. xv, 123 p. \$24.95.

KRUKOWSKI, LUCIAN: Art and Concept: A Philosophical Study. Amherst, MA: University Press, 1987. xi, 127 p. \$20.00.

KUBOVY, MICHAEL: The Psychology of Perspective and Renaissance Art. New York: Cambridge, 1986. 192 p. \$39.50.

KUEHN, MANFRED: Scottish Common Sense in Germany, 1768–1800: A Contribution to the History of Critical Philosophy. Kingston: McGill-Queen's UP, 1987. xiv, 300 p. \$35.00.

KUNO, SUSUMU: Functional Syntax: Anaphora, Discourse and Empathy. Chicago: University Press, 1987. viii, 320 p. Cloth \$49.00, paper \$18.95.

KURSUNOGLU, BEHRAM N., and EUGENE P. WIGNER: Reminiscences about a Great Physicist: Paul Adrien Maurice Dirac. New York: Cambridge, 1987. xviii, 297 p. \$49.50.

- KVART, IGAL: A Theory of Counterfactuals. Indianapolis, IN: Hackett, 1986. xxviii, 283 p. Cloth \$39.50, paper \$19.50.
- LAKOFF, GEORGE: Women, Fire, and Dangerous Things: What Categories Reveal About the Mind. Chicago: University Press, 1987. xvii, 614 p. \$29.95.
- LANZ, PETER: Menschliches Handeln zwischen Kausalität und Rationalität. Frankfurt: Althenaum, 1987. 200 p. 58 DM.
- LARMORE, CHARLES E.: Patterns of Moral Complexity. New York: Cambridge, 1987. Cloth \$32.50, paper \$9.95.
- LATOUR, BRUNO: Science in Action. Cambridge, MA: Harvard, 1987. viii, 274 p. \$25.00.
- LEMAN-STEFANOVIC, INGRID: The Event of Death: A Phenomenological Enquiry. Dordrecht: Nijhoff, 1986. xvi, 345 p. \$66.50.
- LENSON, DAVID: The Birth of Tragedy: A Commentary. Boston: Twayne, 1987. x, 128 p. \$18.95.
- LESSNOFF, MICHAEL: Social Contract. Atlantic Highlands, NJ: Humanities, 1986. x, 178 p. Cloth \$29.95, paper \$8.50.
- LEVINE, ANDREW: The End of the State. London: Verso, 1987. 198 p. Cloth \$39.95, paper \$14.95.
- LEVINS, RICHARD, and RICHARD LEWONTIN: *The Dialectical Biologist*. Cambridge, MA: Harvard, 1985. ix, 303 p. Cloth \$20.00, paper \$8.95.
- LEWIS, DAVID: Philosophical Papers, Vol. II. New York: Oxford, 1987. xvii, 366 p. Cloth \$34.50, paper \$18.95.
- LINDLEY, RICHARD: Autonomy. Atlantic Highlands, NJ: Humanities, 1986. ix, 198 p. Cloth \$29.95, paper \$8.50.
- LINGIS, ALPHONSO: Phenomenological Explanations. Dordrecht: Nijhoff, 1986. x, 112 p. \$29.90.
- LOMASKY, LOREN E.: Persons, Rights, and the Moral Community. New York: Oxford, 1987. ix, 283 p. \$24.95.
- LONGUET-HIGGINS, H. CHRISTOPHER: Mental Processes: Studies in Cognitive Science. Cambridge, MA: MIT Press, 1987. xiii, 424 p. \$35.00.
- LOSEE, JOHN: Philosophy of Science and Historical Enquiry. New York: Oxford, 1987. viii, 153 p. \$32.00.
- LUPER-FOY, STEVEN, ed.: The Possibility of Knowledge: Nozick and his Critics. Totowa, NJ: Rowman & Littlefield, 1987. viii, 335 p. Cloth \$38.50, paper \$16.95.
- LYCAN, WILLIAM G.: Consciousness. Cambridge, MA: MIT Press, 1987. xii, 165 p. \$20.00.
- LYCOS, KIMON: Plato on Justice and Power: Reading Book I of Plato's Republic. Albany: SUNY Press, 1987. Cloth \$39.50, paper \$16.95.
- MACKEY, JAMES P.: Modern Theology: A Sense of Direction. New York: Oxford, 1987. 200 p. Cloth \$26, paper \$8.95.
- MADISON, G. B.: The Logic of Liberty. Westport, CT: Greenwood, 1986. xiv, 293 p. \$37.95.
- MANDELBAUM, MAURICE: Purpose and Necessity in Social Theory. Baltimore, MD: Johns Hopkins, 1987. ix, 197 p. \$25.00.
- MARGOLIS, JOSEPH: Pragmatism Without Foundations. New York: Basil Blackwell, 1987. xix, 320 p. \$34.95.

MARTIN, MICHAEL: The Legal Philosophy of H. L. A. Hart. Philadelphia: Temple, 1987. xi, 315 p. \$34.95.

D. C. P.

MARTIN, ROBERT M.: The Meaning of Language. Cambridge, MA: MIT Press, 1987. ix, 230 p. Cloth \$19.95, paper \$9.95.

матsuo, ноsaku: The Logic of Unity. Albany: SUNY Press, 1987. xvi, 148 p. Cloth \$29.50, paper \$9.95.

McCLOSKEY, MARY A.: Kant's Aesthetic. Albany: SUNY Press, 1987. vi, 184 p. Cloth \$44.50, paper \$16.95.

MELE, ALFRED R.: Irrationality: An Essay on Akrasia, Self-Deception, and Self-Control. New York: Oxford, 1987. x, 183 p. \$24.95.

MERQUIOR, J. G.: Foucault. Berkeley: California UP, 1987. 188 p. Cloth \$25.00, paper \$8.95.

MÉSZÁROS, ISTVÁN: Philosophy, Ideology and Social Science. New York: St. Martin's, 1986. xix, 284 p. Cloth \$29.95, paper \$12.95.

METZGER, HELENE: La Methode philosophique en Histoire des Sciences: Textes 1914-1939. GAD FREUDENTHAL, ed. Paris: Librairie Artheme Fayard, 1987. 277 p.

MILLS, PATRICIA JAGENTOWICZ: Woman, Nature, and Psyche. New Haven, CT: Yale, 1987. xx, 266 p. \$26.50.

MINK, LOUIS O.: Historical Understanding. FAY, BRIAN, EUGENE O. GOLOB, and RICHARD T. VANN, eds. Ithaca, NY: Cornell, 1987. 294 p. \$29.95.

модгак, девоган к.: Aristotle: The Power of Perception. Chicago: University Press, 1987. x, 249 p. \$29.95.

MOEHLE, NATALIA R.: From Myth to Philosophy: Philosophical Implications of the Mythic Understanding of Transtemporal Identity. Lanham, MD: UP of America, 1987. xxii, 134 p. Cloth \$24, paper \$12.25.

MORRIS, THOMAS V.: Anselmian Explorations: Essays in Philosophical Theology. Notre Dame, IN: University Press, 1987. 264 p. \$28.95.

MUNITZ, MILTON K.: Cosmic Understanding. Princeton, NJ: University Press, 1987. ix, 287 p. \$25.00.

MURDOCH, IRIS: Sartre: Romantic Rationalist. New York: Viking, 1987. 158 p. \$14.95.

NEEDHAM, RODNEY: Counterpoints. Berkeley: California UP, 1987. xiv, 251 p. \$30.00.

NEHAMAS, ALEXANDER: Nietzsche: Life as Literature. Cambridge, MA: Harvard, 1987. x, 261 p. Cloth \$20, paper \$8.95.

NELSON, CHARLES H.: John Elof. Boodin. New York: Philosophical Library, 1987. 176 p. \$22.95.

NELSON, PAUL: Narrative and Morality: A Theological Inquiry. University Park: Pennsylvania State UP, 1987. ix, 180 p. \$21.50.

NEVILLE, ROBERT C.: The Puritan Smile: A Look Toward Moral Reflection. Albany: SUNY Press, 1987, 248 p. Cloth \$44.50, paper \$14.50.

NEWMEYER, FREDERICK J.: The Politics of Linguistics. Chicago: University Press, 1987. viii, 172 p. \$23.95.

NICKEL, JAMES W.: Making Sense of Human Rights. Berkeley: California UP, 1987. xiv, 253 p. Cloth \$29.50, paper \$9.95.

- NORMAN, RICHARD: Free and Equal: A Philosophical Examination of Political Values. New York: Oxford, 1987. 178 p. Cloth \$37.00, paper \$12.95.
- O'GRADY, WILLIAM: Principles of Grammar and Learning. Chicago: University Press, 1987. xiii, 233 p. \$27.50.
- OLAFSON, FREDERICK A.: Heidegger and the Philosophy of Mind. New Haven, CT: Yale, 1987. xxi, 294 p. \$23.50.
- OPPENHEIM, FRANK M.: Royce's Mature Philosophy of Religion. Notre Dame, IN: University Press, 1987. xv, 432 p. Cloth \$32.95.
- OSBORNE, CATHERINE: Rethinking Early Greek Philosophy: Hippolytus of Rome and the Presocratics. Ithaca, NY: Cornell, 1987. viii, 383 p. \$47.50.
- PENNER, TERRY: The Ascent from Nominalism. Dordrecht: Reidel, 1987. xxviii, 452 p. \$64.50.
- PEPERZAK, ADRIAAN: Philosophy and Politics: A Commentary on the Preface to Hegel's Philosophy of Right. Dordrecht: Nijhoff, 1986. xi, 144 p. \$40.95.
- PERRETT, ROY W.: Death and Immortality. Dordrecht: Nijhoff, 1986. 202 p. \$38.50.
- PHILLIPS, D. Z.: Belief, Change and Forms of Life. Atlantic Highlands, NJ: Humanities, 1986. xv, 139 p. \$29.95.
- PINKARD, TERRY: Democratic Liberalism and Social Union. Philadelphia: Temple UP, 1987. xix, 219 p. \$24.95.
- POLLNER, MELVIN: Mundane Reason. New York: Cambridge, 1987. xviii, 179 p. \$29.95.
- POST, JOHN F.: The Faces of Existence: An Essay in Nonreductive Metaphysics. Ithaca, NY: Cornell, 1987. 390 p. \$29.95.
- PREUS, J. SAMUEL: Explaining Religion: Criticism and Theory from Bodin to Freud. New Haven, CT: Yale, 1987. xxi, 231 p. \$25.00.
- PRIEST, STEPHEN: Hegel's Critique of Kant. New York: Oxford, 1987. 229 p. \$43.00.
- PRIORESCHI, PLINIO: Man and War. New York: Philosophical Library, 1987. xx, 339 p. \$22.95.
- QUINE, W. V.: Quiddities: An Intermittently Philosophical Dictionary. Cambridge, MA: Harvard, 1987. 249 p. \$20.00.
- REDHEAD, MICHAEL: Incompleteness, Nonlocality, and Realism: A Prolegomenon to the Philosophy of Quantum Mechanics. New York: Oxford, 1987. viii, 191 p. \$45.00.
- REEVE, ANDREW: Property. Atlantic Highlands, NJ: Humanities, 1986. viii, 204 p. Cloth \$29.95, paper \$8.50.
- REGAN, TOM: Bloomsbury's Prophet: G. E. Moore and the Development of his Moral Philosophy. Philadelphia: Temple, 1987. xix, 307 p. \$34.95.
- REGVALD, RICHARD: Heidegger et le Probleme du Néant. Dordrecht: Nijhoff, 1986. xiii, 201 p. \$54.00.
- REINES, ALVIN J.: Polydoxy: Explorations in a Philosophy of Liberal Religion. Buffalo, NY: Prometheus, 1987. 219 p. \$22.95.
- RESCHER, NICHOLAS: Ethical Idealism: An Inquiry into the Nature and Function of Ideals. Berkeley: California UP, 1987. xi, 148 p. \$30.00.
- RESNICK, STEPHEN A., and RICHARD D. WOLFF: Knowledge and Class. Chicago: University Press, 1987. vii, 352 p. \$32.50.

- Digitized by Arya Samaj Foundation Chennai and eGangotri
- RESNIK, MICHAEL D.: Choices: An Introduction to Decision Theory. Minneapolis: Minnesota UP, 1987. xiii, 221 p. Cloth \$29.50, paper \$12.95.

The state of

- RICHARDSON, JOHN: Existential Epistemology: A Heideggerian Critique of the Cartesian Project. New York: Oxford, 1987. ix, 210 p. \$39.95.
- RICKERT, HEINRICH: The Limits of Concept Formation in Natural Science. New York: Cambridge, 1986. xxxii, 240 p. Cloth \$49.50, paper \$14.95.
- RIGBY, S. H.: Marxism and History: A Critical Introduction. New York: St. Martin's, 1987. vi, 314 p. \$37.50.
- ROBERTS, DAVID D.: Benedetto Croce and the Uses of Historicism. Berkeley: California UP, 1987. xii, 449 p. \$30.00.
- ROHRLICH, FRITZ: From Paradox to Reality: Our Basic Concepts of the Physical World. New York: Cambridge, 1987. x, 227 p. \$34.50.
- ROLSTON, HOLMES III: Science and Religion: A Critical Survey. Philadelphia: Temple UP, 1987. x, 358 p. \$34.95.
- ROSENBLUM, NANCY L.: Another Liberalism: Romanticism and the Reconstruction of Liberal Thought. Cambridge, MA: Harvard, 1987. 225 p. \$25.00.
- ROSENTHAL, ABIGAIL L.: A Good Look at Evil. Philadelphia: Temple UP, 1987. xiv, 245 p. \$24.95.
- ROTH, PAUL A.: Meaning and Method in the Social Sciences: A Case for Methodological Pluralism. Ithaca, NY: Cornell, 1987. x, 250 p. \$26.95.
- ROUSE, JOSEPH: Knowledge and Power: Toward a Political Philosophy of Science. Ithaca, NY: Cornell, 1987. xvi, 283 p. \$29.95.
- RUNDELL, JOHN F.: Origins of Modernity. Madison: Wisconsin UP, 1987. vi, 249 p. \$36.00.
- RUSSMAN, THOMAS A.: A Prospectus for the Triumph of Realism. Macon, GA: Mercer UP, 1987. viii, 157 p. \$24.95.
- SAFFORD, JOHN LUGTON: Pragmatism and the Progressive Movement in the United States: The Origin of the New Social Sciences. Lanham, MD: UP of America, 1987. xiii, 255 p. Cloth \$26.50, paper \$13.25.
- SALLIS, JOHN: Spacings: Of Reason and Imagination in Texts of Kant, Fichte, Hegel. Chicago: University Press, 1987. xvi, 177 p. Cloth \$25.00, paper \$10.95.
- SAPONTZIS, S. F.: Morals, Reason, and Animals. Philadelphia: Temple UP, 1987. xix, 302 p. \$34.95.
- SARTORI, GIOVANNI: The Theory of Democracy Revisited. Vol. I: The Contemporary Debate; Vol. II: The Classical Issues. Chatham, NJ: Chatham House, 1987. 542 p. 2-vol. set: cloth \$40.00, paper \$21.95.
- SATRIS, STEPHEN: Ethical Emotivism. Dordrecht: Nijhoff, 1986. xii, 201 p. \$48.50.
- SCHELER, MAN: Person and Self-Value. Dordrecht: Nijhoff, 1987. xxix, 201 p. \$59.50.
- schiffer, stephen: Remnants of Meaning. Cambridge, MA: MIT Press, 1987. xx, 303 p. \$27.50.
- SCHMITT, RICHARD: Introduction to Marx and Engels: A Critical Reconstruction. Boulder, CO: Westview Press, 1987. xviii, 220 p. Cloth \$38.50, paper \$15.95.

SCHOR, NAOMI: Reading in Detail: Aesthetics and the Feminine. New York: Methuen, 1987. x, 184 p. Cloth \$25.00, paper \$10.95.

SCOTT, CHARLES E.: The Language of Difference. Atlantic Highlands, NJ: Humanities, 1987. viii, 188 p. \$35.00.

SEN, AMARTYA: *The Standard of Living*. New York: Cambridge, 1987. xiv, 125 p. \$24.95.

SHANKER, S. G.: Wittgenstein and the Turning-Point in the Philosophy of Mathematics. Albany: SUNY Press, 1987. xi, 358 p. Cloth \$44.50, paper \$14.95.

SHEEHAN, THOMAS: Karl Rahner: The Philosophical Foundations. Athens: Ohio UP, 1987. xii, 320 p. \$24.95.

SHER, GEORGE: Desert. Princeton, NJ: University Press, 1987. xiii, 215 p. \$28.50.

SINGER, IRVING: The Nature of Love, Vol. 3: The Modern World. Chicago: University Press, 1987. xiii, 473 p. \$24.95.

SLESS, DAVID: In Search of Semiotics. Totowa, NJ: Barnes & Noble, 1986. 170 p. \$28.50.

SMART, J. J. C.: Essays Metaphysical & Moral. New York: Basil Blackwell, 1987. viii, 312 p. \$49.95.

SOFFER, WALTER: From Science to Subjectivity: An Interpretation of Descartes' Meditations. New York: Greenwood, 1987. xv, 183 p. \$35.00.

SOLOMON, ROBERT C.: From Hegel to Existentialism. New York: Oxford, 1987. xii, 299 p. \$29.95.

SOLTAN, KAROL EDWARD: The Causal Theory of Justice. Berkeley: California UP, 1987. xii, 265 p. \$30.00

SPRUNG, MERVYN: The Magic of Unknowing: An East-West Soliloquy. Lewiston, NY: Broadview, 1987. 159 p. \$15.95.

STADLER, INGRID: Contemporary Art and Its Philosophical Problems. Buffalo, NY: Prometheus, 1987. 153 p. \$21.95.

STAFFORD, WILLIAM: Socialism, Radicalism and Nostalgia: Social Criticism in Britain, 1775–1830. New York: Cambridge, 1987. viii, 304 p. Cloth \$42.50, paper \$15.95.

STRAWSON, GALEN: Freedom and Belief. New York: Oxford, 1987. vii, 339 p. \$39.95.

STROKER, ELISABETH: Investigations in Philosophy of Space. Athens: Ohio UP, 1987. viii, 321 p. \$28.95.

SUMMERS, DAVID: The Judgment of Sense: Renaissance Naturalism and the Rise of Aesthetics. New York: Cambridge, 1987. xiii, 365 p. \$39.50.

TAMÁG, GYÖRGY: The Logic of Categories. ROBERT S. COHEN, ed. Dordrecht: Reidel, 1987. xiv, 524 p. \$89.50.

TAYLOR, MICHAEL: The Possibility of Cooperation. New York: Cambridge, 1987. xiii, 205 p. Cloth \$39.50, paper \$12.50.

TEN, C. L.: Crime, Guilt, and Punishment. New York: Oxford, 1987. 175 p. Cloth \$39.95, paper \$18.95.

TENNANT, NEIL: Anti-Realism and Logic: Truth as Eternal. New York: Oxford, 1987. xii, 325 p. \$64.00.

THOMPSON, ROBERT A., LOUISE S. THOMPSON: Egoshell. Buffalo, NY: Prometheus, 1987. 369 p. \$22.95.

- TRACY, DAVID: Plurality and Ambiguity. San Francisco: Harper and Row, 1987. xii, 148 p. \$15.95.
- TURNER, MARK: Death is the Mother of Beauty. Chicago: University Press, 1987. xi, 208 p. \$16.95.
- TURSMAN, RICHARD: Peirce's Theory of Scientific Discovery. Bloomington: Indiana UP, 1987. xi, 160 p. \$25.00.
- VOEGELIN, ERIC: In Search of Order. Baton Rouge: Louisiana State UP, 1987. 120 p. \$14.95.
- VON MOLNÁR, GÉZA: Romantic Vision, Ethical Context: Novalis and Artistic Autonomy. Minneapolis: Minnesota UP, 1987. xxxiv, 246 p. Cloth \$35.00, paper \$14.95.
- WALICKI, ANDRZEJ: Legal Philosophies of Russian Liberalism. New York: Oxford, 1987. ix, 477 p. \$86.00.
- walzer, Michael: Interpretation and Social Criticism. Cambridge, MA: Harvard, 1987. viii, 96 p. \$12.50.
- WANG, HAO: Reflections on Kurt Gödel. Cambridge, MA: MIT Press, 1987. xxvi, 336 p. \$25.00.
- WARNER, RICHARD: Freedom, Enjoyment, and Happiness: An Essay on Moral Psychology. Ithaca, NY: Cornell, 1987. 181 p. \$19.95.
- WARNKE, GEORGIA: Gadamer: Hermeneutics, Tradition, and Reason. Stanford, CA: University Press, 1987. xi, 206 p. Cloth \$35.00, paper \$11.95.
- WATSON, RICHARD A.: The Breakdown of Cartesian Metaphysics. Atlantic Highlands, NJ: Humanities Press, 1987. xii, 240 p. \$45.00.
- WEINHEIMER, HEINZ: Rationalität und Bergründung. Bonn: Bouvier, 1986. 213 p. 58 DM.
- WEINREB, LLOYD L.: Natural Law and Justice. Cambridge, MA: Harvard, 1987. ix, 320 p. \$25.00.
- weissman, david: Intuition and Ideality. Albany: SUNY Press, 1987. ix, 309 p. Cloth \$39.50, paper \$14.95.
- WELBOURNE, MICHAEL: The Community of Knowledge. Atlantic Highlands, NJ: Humanities, 1986. xiv, 100 p. Cloth \$25.00, paper \$9.95.
- wilcox, David J.: The Measure of Times Past: Pre-Newtonian Chronologies and the Rhetoric of Relative Time. Chicago: University Press, 1987. ix, 292 p. \$27.50.
- WRIGHT, CRISPIN: Realism, Meaning and Truth. New York: Basil Blackwell, 1987. xii, 386 p. \$39.95.
- WUTHNOW, ROBERT: Meaning and Moral Order: Explorations in Cultural Analysis. Berkeley: California UP, 1987. xiii, 435 p. \$38.50.
- YOUNG, JULIAN: Willing and Unwilling: A Study in the Philosophy of Arthur Schopenhauer. Dordrecht: Nijhoff, 1987. xiii, 169 p. \$39.50.
- YOUNG, J. Z.: Philosophy and the Brain. New York: Oxford, 1987. viii, 233 p. \$22.50.
- YOUNG, PAUL: The Nature of Information. New York: Praeger, 1987. x, 192 p. \$35.00.

NEW BOOKS IN PAPERBACK

Digitized by Arya Samaj Foundation Chennai and eGangotri

NEW BOOKS IN PAPERBACK

ALMOND, BRENDA: Moral Concerns. Atlantic Highlands, NJ: Humanities, 1987. xi, 152 p. \$9.95.

BARRETT, WILLIAM: Death of the Soul: From Descartes to the Computer. New York: Doubleday, 1987. xvi, 173 p. \$8.95.

BONJOUR, LAURENCE: The Structure of Empirical Knowledge. Cambridge, MA: Harvard, 1987. xiii, 285 p. \$10.95.

BRANDL, JOHANNES: Die Notwendigkeit der Namen: Über das Benennen von Einzeldingen und die Relevanz einer historischekausalen Grundlagen für den Aufbau einer allgemeinen Theorie der Bedeutung. Graz: Technische Universität Verlag, 1987. vii, 509 p. 36 DM.

CONWAY, DAVID: A Farewell to Marx. New York: Penguin, 1987. 231 p. \$6.95.

DANCY, JONATHAN: Berkeley: An Introduction. New York: Basil Blackwell, 1987. 165 p. \$12.95.

DEN UYL, DOUGLAS J., and DOUGLAS B. RASMUSSEN, eds.: The Philosophic Thought of Ayn Rand. Champaign: Illinois UP, 1987. \$12.95.

DEVITT, MICHAEL, and KIM STERELNY: Language and Reality: An Introduction to the Philosophy of Language. Cambridge, MA: MIT Press, 1987. xii, 274 p. \$12.50.

DIPPEL, STEWART A.: A Study of Religious Thought at Oxford and Cambridge: 1590-1640. Lanham, MD: UP of America, 1987. xiii, 122 p. \$9.75.

DUMMETT, MICHAEL: The Interpretation of Frege's Philosophy. Cambridge, MA: Harvard, 1987. \$17.50.

DUNN, ROBERT: The Possibility of Weakness of Will. Indianapolis, IN: Hackett, 1987. x, 175 p. \$14.50.

FEINBERG, JOEL: Harm to Others: The Moral Limits of the Criminal Law. New York: Oxford, 1987. xiii, 269 p. \$9.95.

FOGELIN, ROBERT J.: Wittgenstein. 2nd ed. New York: Routledge & Kegan Paul, 1987. xv, 255 p. \$12.95.

GALISON, PETER: How Experiments End. Chicago, University Press, 1987. xii, 330 p. \$14.95.

GAUTHIER, DAVID: Morals by Agreement. New York: Oxford, 1987. vi, 367 p. \$13.95.

GOODMAN, NELSON: Of Mind and Other Matters. Cambridge, MA: Harvard, 1987. 210 p. \$8.95.

GROTH, MILES: Preparatory Thinking of Heidegger's Teaching. New York: Philosophical Library, 1987. 159 p. \$9.95.

GUTTENPLAN, SAMUEL: The Language of Logic. New York: Basil Blackwell, 1987. x, 336 p. \$16.95.

HAMPSHIRE, STUART: Morality and Conflict. Cambridge, MA: Harvard, 1983. 175 p. \$9.50.

HASTINGS CENTER: Guidelines on the Termination of Life-Sustaining Treatment and the Care of the Dying. New York: Hastings Center, 1987. xii, 159 p. \$14.50.

JOHNSON, ROBERT M.: A Logic Book. Belmont, CA: Wadsworth, 1987. xiv, 271 p. \$17.75.

a fine

JOLLEY, NICHOLAS: Leibniz and Locke: A Study of the New Essays on Human Understanding. New York: Oxford, 1987. viii, 215 p. \$17.95.

JORDAN, JAMES N.: Western Philosophy: From Antiquity to the Middle Ages.

New York: Macmillan, 1987. xvi, 492 p. \$15.00.

KLUBACK, WILLIAM: The Idea of Humanity: Hermann Cohen's Legacy to Phil-

osophy and Theology. Lanham, MD: UP of America, 1987. viii, 304 p. \$15.75.

KRAUT, RICHARD: Socrates and the State. Princeton, NJ: University Press, 1987. ix, 338 p. \$10.50.

LAFRANCE, GUY: Pouvoir et tyrannie. Ottawa: University Press, 1986. 181 p. \$19.95.

LO, P. C.: Treating Persons as Ends: An Essay on Kant's Moral Philosophy. Lanham, MD: UP of America, 1987. xix, 347 p. \$16.75.

MANOLIDIS, GEORGIOS: Die Rolle der Physiologie in der Philosophie Epikurs. Frankfurt: Athenäum, 1987. 188 p. 34 DM.

MARTINEAU, ALAIN: Herbert Marcuse's Utopia. Montreal: Harvest House, 1986. 156 p. \$9.95.

MILLER, GEORGE A., and PHILIP N. JOHNSON-LAIRD: Language and Perception. Cambridge, MA: Harvard, 1976. viii, 766 p. \$12.50.

NAGEL, THOMAS: What Does It All Mean?: A very Short Introduction to Philosophy. New York: Oxford, 1987. 101 p. \$12.95.

OBSON, JOHN M., ed.: Origin and Evolution of the Universe: Evidence for Design? Montreal: McGill-Queen's UP, 1987. xiii. 297 p. \$15.95.

PETERSON, FORREST H.: John Dewey's Reconstruction in Philosophy. New York: Philosophical Library, 1987. 95 p. \$8.50.

POLKINGHORNE, JOHN: One World: The Interaction of Science and Theology. Princeton, NJ: University Press, 1987. xiii, 114 p. \$7.95.

REGIS, EDWARD JR., ed.: Extraterrestrials: Science and Alien Intelligence. New York: Cambridge, 1987. x, 278 p. \$12.95.

RICHARDS, STEWART: Philosophy and Sociology of Science. New York: Basil Blackwell, 1987. x, 241 p. \$15.95.

RIGBY, PAUL: Original Sin in Augustine's Confessions. Ottawa: University Press, 1987. ix, 137 p. \$17.95 (CAN).

RING, MERRILL: Beginning with the PreSocratics. Palo Alto, CA: Mayfield, 1987. xii, 166 p. \$9.95.

ROLSTON, III, HOLMES: Science and Religion: A Critical Survey. New York: Random House, 1987. viii, 358 p. \$11.50.

RUF, HENRY L.: Investigating Philosophy. Lanham, MD: UP of America, 1987. xii, 492 p. \$23.50.

RUSE, MICHAEL: Taking Darwin Seriously. New York: Basil Blackwell, 1987. xv, 303 p. \$15.95.

RUSSELL, BERTRAND: Bertrand Russell on Ethics, Sex, and Marriage. AL SECKEL, ed. Buffalo, NY: Prometheus, 1987. vii, 348 p. \$14.95.

SAYRE, KENNETH M.: Plato's Late Ontology: A Riddle Resolved. Princeton, NJ: University Press, 1983. x, 328 p. \$16.50.

SCHWEITZER, ALBERT: The Philosophy of Civilization. Buffalo, NY: Prometheus, 1987. vii, 347 p. \$13.95.

SMITH, WILFRED CANTWELL: Faith and Belief. Princeton, NJ: University Press, 1987. ix, 347 p. \$12.50.

SPRINKER, MICHAEL: Imaginary Relations. New York: Methuen, 1987. 302 p. \$14.95.

THOMAS, JOHN E., and WILFRID J. WALUCHOW: Well and Good: Case Studies in Biomedical Ethics. Lewiston, NY: Broadview, 1987. 192 p. \$10.95.

UNGER, ROBERTO MANGABEIRA: False Necessity: Anti-Necessitarian Social Theory in the Service of Radical Democracy. New York: Cambridge, 1987. xiii, 653 p. \$59.50; Plasticity Into Power: Comparative-Historical Studies on the Institutional Conditions of Economic and Military Success. New York: Cambridge, 1987. vi, 231 p. \$37.50; Social Theory: Its Situation and Its Task. New York: Cambridge, 1987. vii, 256 p. \$39.50.

URBACH, EPHRAIM E.: The Sages: Their Concepts and Beliefs. Cambridge, MA: Harvard, 1987. xx, 1084 p. \$18.95.

URBACH, PETER: Francis Bacon's Philosophy of Science: An Account and a Reappraisal. La Salle, IL: Open Court, 1987. 209 p. \$9.95.

VELASQUEZ, MANUEL G.: Business Ethics: Concepts and Cases. 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1988. xii, 420 p. \$28.67.

WENDEL, HANS JÜRGEN: Benennung, Sinn, Notwendigkeit. Frankfurt: Athenäum, 1987. 240 p. 38 DM.

WOLGAST, ELIZABETH H.: The Grammar of Justice. Ithaca, NY: Cornell, 1987. xiii, 219 p. \$9.95.

Semiotics from Peirce to Barthes

A Conceptual Introduction to the Study of Communication, Interpretation and Expression

by

V. TEJERA

1988. (ix, 201 p.) ISBN 90 04 08597 1

ca. Gld. 72.-/US\$ 36.-

The synoptic nature of this book, the level of clarity at which it is pitched, makes it particularly useful to students of literature and philosophy looking for an introduction to semiotics:

Order from your regular supplier or from:



E. J. BRILL P.O. Box 9000 2300 PA Leiden The Netherlands

THE PHILOSOPHICAL QUARTERLY

Editorial Chairman: Neil Cooper

Founded by the Scots Philosophical Club and the University of St Andrews in 1950, **The Philosophical Quarterly** aims to foster and publish significant contributions in every branch of the subject, promoting discussion of recent philosophical works, and providing critical reviews of recent books, in a style accessible to as wide a group of philosophers as possible.

Published January, April, July, October for the Scots Philosophical Club and the University of St Andrews

Subscription Rates Volume 37, 1987

Individuals: £11 (UK) £13.75 (overseas) US\$26 (N. America+Japan) Institutions: £27.50 (UK) £36.75 (overseas) US\$61.50 (N. America+Japan)

Rates include inland or accelerated surface postage.

Please send orders with payment to:

Iris Taylor Journals Department Basil Blackwell 108 Cowley Road Oxford OX4 1JF England or to: Journals Department Basil Blackwell PO Box 1320 Murray Hill Station NY 10156 USA

Basil Blackwell · Oxford and New York

SOCIAL PHILOSOPHY AND POLICY

Edited by Ellen Frankel Paul

Each 192-page issue of **Social Philosophy and Policy** is devoted to a single topic in social or political philosophy, with specially invited papers (and, for designated issues, unsolicited papers) by leading scholars on the most important issues of our time. The topics for Volume 4 are **Philosophy and Law**, and **Philanthropy**. Previous issues have focused on **Distributive Justice**, **Human Rights**, **Liberty and Equality**, **Ethics and Economics**, **Nuclear Rights/Nuclear Wrongs**, and **Marxism and Liberalism**, with contributions from such distinguished philosophers as Maurice Cranston, Ronald Dworkin, R. M. Hare, David Gauthier, J. R. Lucas, Amartya Sen, Jon Elster, and G. A. Cohen, among others.

October and April for the Social Philosophy and Policy Center, Bowling Green State University, USA

Subscription to Volume 4, 1986/7 Individuals: £11 (UK) £12.50 (overseas) US\$15 (N. America & Japan) Institutions: £26 (UK) £31.50 (overseas) US\$45 (N. America & Japan) Please send orders with payment to:

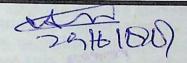
Iris Taylor Journals Department Basii Blackwell 108 Cowley Road Oxford OX4 1JF England

or to: Journals Department Basil Blackwell PO Box 1320 Murray Hill Station NY 10156 USA

Basil Blackwell · Oxford and New York

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 3, MARCH 1988



DUCKING HARM*

FTEN a person A can transfer a threatened harm to another person B. Often too there are two ways A can achieve this result, one more acceptable than the other. One way is for A merely to exit the threatening situation, leaving B behind to suffer the injury. The other way is for A directly to place B in harm's way as A's shield. This distinction between ducking harm to B and using B as a shield—or, as we shall often call it, "sacrificing" B—is neglected in law and moral philosophy. It is an analogue to the act/omission distinction, arousing similar intuitions, but wholly within the realm of acts. If accepted, it is a new nonconsequentialist constraint on distributing harm, and its explanation may be an adequacy condition on solutions to the act/omission problem.

In a recent joke two campers, Alex and Bruce, meet a ravenous bear. As Alex grabs his running shoes, Bruce points out that no one can outrun a bear. "I don't have to outrun him," Alex replies. "I only have to outrun you." Few contemporary Westerners³ will criticize

* We thank Julia Driver, Leslie Friedman Goldstein, David Haslett, Douglas Stalker, and Roslyn Weiss for helpful conversations, and Paul Robinson and Stephen Schulhofer for advice on law. We do not duck final responsibility.

We regret our lack of a term as simple and natural as 'duck' for the other side of the distinction. 'Enshield' is accurate but artificial; as noted below, it fails in scarce-resource cases. The lack of a natural term may suggest the existence of several subdistinctions that more detailed analysis would reveal.

² Ducking is an occasional literary theme. Canadian novelist Robertson Davies' Deptford trilogy (Fifth Business, The Manticore, and World of Wonders) deals at length with the psychological effects of a childhood ducking incident on Dunstan Ramsay. In Orwell's 1984, Winston Smith's final degradation, as he faces the rat torture, is his cry: "Do it to Julia!"

³ Possibly in other times and cultures, prevailing ethics has condemned ducking harm as not only unheroic, but impermissible. Medieval chivalry is one Western tradition where some heroic self-sacrifice was obligatory, especially when A was a knight and B a woman or child. Similar attitudes characterized feudal Japan and

0022-362X/88/8503/0115\$02.00

© 1988 The Journal of Philosophy, Inc.

Alex for running away full tilt, or even for using his new Sauconys. But suppose Alex instead ties Bruce's ankles, or knocks Bruce unconscious and throws him to the bear. Alex is now blameworthy in ethics and in law. The result is the same in both cases: Bruce's death. Further, Alex may know the result to be the same if he knows he can outrun Bruce. Nonetheless most people sharply distinguish the two acts.

The bear example has unusual features—e.g., that Alex survives by Bruce's death—and so we will usually discuss purer examples of ducking. A standard pure case is this one:

MALL GUNMAN:

- (a1) Angela, at the end of a movie ticket line, sees X about to shoot a .22 automatic at her. Angela knows that a .22 bullet will kill one person but not two. Angela leaps aside; the bullet kills Brenda, who is next in line.
- (a2) Same as (a1), but Angela grabs Brenda and moves her in front as a shield; the bullet kills Brenda.

A slight variant which is useful later is a line of cars:

SPEEDING TRUCK:

- (b1) Arthur, at the end of a line of stopped traffic, sees a runaway truck in his rearview mirror. Arthur changes lanes; the truck crashes into Brian's car, injuring him.
- (b2) Same as (b1), but Arthur beckons to a new driver, Brian, to join the line behind him; Brian does so and is injured by the truck.

Other situations offer more subtle examples of the duck/sacrifice distinction.

TERRORISTS:

- (c1) Alison is one of 25 U.S. government officials on an airplane, each with a briefcase bearing an official seal. Terrorist hijackers announce they will kill one American per hour until their demands are met. Surreptitiously Alison covers her seal with a Libya Air sticker. The terrorists pass her briefcase and shoot Beatrice, the next American.
- (c2) Same as (c1), but Alison has no Libyan sticker. Instead she switches briefcases with Babette, a French novelist, while she is in the bathroom. The terrorists shoot Babette.

ancient Rome, to name only two. The orderly and gracious self-sacrifice of male *Titanic* passengers will be long remembered.

For good or ill, contemporary support for the duck/sacrifice distinction is strong. Of 283 Delaware ethics students, before discussion, 84 per cent said that ducking the bullet or the bear was not immoral and that the sacrificial alternative was immoral. If we exclude those who gave different verdicts on ducking bullets and bears, support for the distinction rises to 91 per cent.

Such cases raise moral and legal puzzles. The moral puzzle, like that posed by parallel acts and omissions, is to explain why we are strongly inclined to view A as blameless in the 1 but not in the 2 cases. At a minimum, we find A less blameworthy in 1. What morally relevant difference is there between ducking and forced shielding, when both are acts of A that A knows will lead to B's death? Similarly, the legal puzzle is what principles can assign A lower civil and criminal liability for ducking. What makes the legal puzzle especially acute, however, is the remarkable fact that, on standard accounts of homicide, A seems to have no defense to murder or manslaughter even for ducking. This surprising claim is explained in section III.

There is no doubt that ducking is approved in a wide range of cases. It is usually conceded that anyone has the right to escape disaster even when not everyone can escape. One may rush to the exit of a burning theater, though one will be criticized for trampling others in one's flight. It is equally clear to most people that forced shielding is wrong, and felons are often convicted of murder for the death of their human shields. B's self-defense rights are symmetrical: B may forcibly resist being sacrificed, but he may not prevent A from ducking, since that would be to sacrifice A. Likewise, a similar distinction can be drawn in aggressive competition for scarce resources. One may rush to buy food in a famine or to sell stock on bad news. In a variant of the camper tale, Alex races Bruce for a single dose of rattlesnake antidote; the shift from escaping a threat to securing a benefit has no effect on intuition.

An important difference is that in scarce-resource situations, "sacrificing" B is not enshielding B, but merely wrongful direct interference with him. The previous examples illustrate this point, as well as the following case:

SINKING BOATS:

- (d1) Arnold's only escape from his sinking boat is a rescue helicopter equidistant from him and Ben's sinking boat. Given the fog, the rescuers will follow the stronger distress signal. Arnold improvises a large aerial to strengthen his signal; he is saved and Ben drowns.
- (d2) Same as (d1), but Arnold makes his signal stronger by jamming Ben's signal; the rescuers are diverted to Arnold.

⁴ Wayne R. LaFave and Austin W. Scott, Jr., Handbook on Criminal Law (St. Paul, MN: West, 1972), p. 550. Their explanation: "When the robber intentionally places the victim in danger of being shot, the latter's death is clearly the foreseeable consequence of the former's actions." Such convictions may be had on ordinary causal and mens rea principles alone, without the felony-murder doctrine; e.g., R. v. Pagett (London Times, Feb. 4, 1983) (manslaughter); Pizano v. Superior Court of Tulare County, 577 P.2d 659 (1978) (murder).

Because sacrificing here lacks the element of enshielding, the duck/sacrifice distinction may be different or less clear. Our discussion will concentrate on the active-threat type (a)-(c); but the scarce-resource variety should be kept in mind, especially regarding solutions that employ the concept of using B as a means.

It is important to see that the ducking puzzle cannot be solved by mere appeal to "self-defense." Self-defense cannot distinguish ducking from sacrificing; the puzzle is how to distinguish among acts of self-defense. In case (a), A escapes the gunman's bullet exactly as much by ducking as by using B as a shield, and both the situation and the effect on B are the same. Legally, none of the acts falls under self-defense anyway, since the death caused is of B, an innocent bystander. Many ethical writers on abortion endorse killing the innocent in self-defense. Often what they endorse, however, is killing an innocent "assailant," whereas in our examples the innocent victim is not the threat. Among other objections, a blanket endorsement of all killing to save one's life violates the widespread expectation that people on waiting lists for organ transplants will let themselves die rather than remove rivals higher on the list, or kill healthy persons to expand the organ pool.

Also, the moral and legal issues are not restricted to defense of self or defense of body. The same puzzle arises for non-self victims and non-death harms. With the mall gunman, Y may be standing outside the line and save A either by pushing A aside or by pushing B in front of A. Morally speaking, Y is likely to be criticized for pushing B into the path of a bullet but not for pushing A out of it. Legally, it seems obvious that pushing A out of death's path should not subject Y to criminal charges or civil liability to B's relatives. The puzzle also survives changing the harm from death to bodily injury or property damage. Doubtless A may wall in his property against a coming flood, even if the water is thereby diverted to the home of B; but he may not build his wall out of the bricks of B's doomed home. In an era of faculty cutbacks, one may save one's job by writing faster, but not by burning one's colleagues' manuscripts, even if they are worthless.

II.

At least seven factors can change our evaluation of the morality of ducking. At first sight, each alters either the absolute acceptability of ducking or its comparative acceptability with sacrificing.

(i) Responsibility for threat. A may lose ducking rights to the extent that he created the danger or has complicity in its creation. A's duck is clearly wrong if he hired the gunman or rented the bear.⁵

⁵ These acts make A guilty of murder if B dies, and they justify B's sacrificing A in self-defense, as Marlowe does at the end of the movie *The Big Sleep*.

If A is only accidentally or negligently, not intentionally, responsible for the threat, his rights are less clear. How much does it matter if the bear was lured by A's odor or that of the food cooler A forgot to close, or if the enraged gunman is A's dismissed employee?

(ii) Innocence of act. Presumably A's duck is less acceptable the more it involves intrinsically wrong acts. A is more blameworthy if he can escape only by lying or stealing than if he merely runs away. This factor is considered in section IV as a solution to the basic puzzle.

(iii) Equal quantity of harm. A's duck is worse insofar as the injury A avoids is less than the injury B suffers. In (b1), if A's vehicle is a heavy truck and B's a small car, surely A must not prefer B's death to his own property damage.

(iv) Equal probability of harm. A is less justified in ducking insofar as he is less likely to be injured than B. In (a1), if A has an invisible bulletproof vest, his fear of a head shot may not justify him in ducking the bullet to unprotected B.

- (v) Benefit or cost to A. A may have more right to duck if he has paid a prior cost for being able to do so. This idea explains the morality of queues when those first in line have waited the longest. Similarly, contributors to blood banks win priority in a blood shortage. Conversely, ducking is criticized when the ducker benefits from the harm's befalling someone (i.e., benefits beyond avoiding it himself). Thus the strong denunciation of wartime draft dodgers: as citizens they unfairly benefit from others' injuries in saving the common nation.
- (vi) Voluntariness. The more voluntarily B assumes risk, the clearer it is that A may duck the risked harm to B. Drivers in fog are shamelessly guided by the lights of more reckless drivers who pass them. A feels better running from the bear on a safari than on a camping trip.
- (vii) *Identifiability of victim*. Another familiar factor from the act/omission debate is whether the victim of a duck can be named in advance. The fleeing resistance fighter feels less guilt knowing the Gestapo will arrest someone else than knowing it will arrest his housemate. In the famous "squib" case, 6 one prefers throwing the

The point of this case is that Shepherd is not excused from causing the harm, since the acts of Willis and Ryal were normal reactions. Perkins and Boyce write: "It is a normal human response for one threatened by harmful force, deadly or non-

⁶ Scott v. Shepherd, 2 Wm. Blackstone 892, 96 Eng. Rep. 525 (1773). "Shepherd threw a lighted squib into a crowded marketplace and it fell near Willis who, to protect himself, threw it away and it landed near Ryal who also threw it for his protection. It then fell upon Scott and exploded at that instant putting out one of his eyes. It was held that Shepherd had put out Scott's eye" [Rollin M. Perkins and Ronald N. Boyce, Criminal Law (Mineola, NY: Foundation Press, 1982), p. 809].

lighted firecracker randomly into the crowd to placing it in a specific

person's lap.

Although these factors affect our feelings about the morality of ducking, not all survive analysis. Factors (iii) and (iv), quantity and probability of harm, have an easy utilitarian justification. But it is hard to see factor (vii), identifiability of victim, as serious ethics. And, if the right to a pure duck is clear, factors (v) and (vi) cannot make duckers more justified, only more comfortable, in exercising it.

To focus moral discussion further, we now analyze ducking according to standard legal principles.7 To avoid confusion by the gunman's guilt, let us consider (a1) with a nonhuman danger: say, a boulder rolling with just enough momentum to crush one person. A's leap is a voluntary act. A thus satisfies the first legal requirement for a crime: a voluntary act or omission where there is a special duty to act. A's leap also satisfies both prongs of the standard test for legal causation of B's death. First, A's leap is a necessary condition of B's death ("cause in fact"), since, if A remains, A dies and B lives. Second, no unforeseeable intervening cause deprives A's act of legal responsibility. No cause temporally intervenes, since the boulder was in motion before A leaped; furthermore, A foresees the chain of causation before acting. Thus, A's leap is a legal cause of B's death.8

deadly, to take action in the effort to avoid the impending harm." Because Willis and Ryal acted predictably, their acts are not superseding causes, and Shepherd remains responsible. Our question is: Why are not Willis and Ryal also guilty? To call their behavior normal does not explain their defense. A common-sense causal chain can run through a foreseeable criminal act, as in soliciting crime or lying to a jealous husband. Then both parties are guilty, although the law calls the procurer's guilt

complicity rather than causation.

⁷ For standard expositions of the concepts in this section, see LaFave and Scott, op. cit. (2e 1986); Perkins and Boyce, op. cit.; Jerome Hall, General Principles of Criminal Law (New York: Bobbs-Merrill, 2nd ed., 1960); Glanville Williams, Criminal Law: The General Part (London: Stevens, 2nd ed., 1961); George Fletcher, Rethinking Criminal Law (Boston: Little, Brown, 1978); and H. L. A. Hart and A. M. Honore, Causation in the Law (New York: Oxford, 2nd ed., 1985). None of these sources discusses ducking in "self-defense"; so perhaps no Anglo-American defendant has ever been charged with ducking. But for the most part unblockedhazard murders are not discussed either.

⁸ A Model Penal Code rule on legal cause finds no causation if "the result was too remote or too accidental in its occurrence to have a [just] bearing on the actor's liability or on the gravity of his offense" [2.03.3(b)]. As a full test of causation, this "rule" (an appeal to jury conscience) might be vague enough for a duck to fail to cause death. But in the MPC it is only an exception to an exception, for harm of a foreseen type caused in an unforeseen way. The Explanatory Comment (p. 262) reiterates the standard rule that one is liable for completely foreseen results, i.e., when "events transpire as the actor intended or knew that they would." That is the situation in ducking. See also Hart and Honore, op. cit., pp. 170-172, on the maxim "Intended consequences can never be too remote."

Finally, because A foresees this result, he intends it in law. That is, in a philosophically preferable formulation, A has the state of intent or knowledge that suffices for murder.

These points suggest that A by leaping knowingly kills B, and, therefore, A is guilty of homicide unless A has a recognized defense. Self-defense covers injury only to assailants, whereas B is an innocent backstander. In theory the defense of necessity applies to any choice of a lesser evil under constraint of natural forces. It is here unavailable to A because A does not choose a lesser evil, merely substituting one death for another. Duress, the only other relevant possibility, is traditionally restricted to human threats, with natural dangers left to necessity. Also, duress is traditionally held not to excuse homicide. On this description, duress is doubly inapplicable to A, since the boulder is a natural evil and A's act is homicide.

In sum, A's duck, a voluntary act, is a known necessary condition of B's death, with no unforeseen intervening causes; and no classical defense seems to apply. If A is to escape criminal liability, legal theory must somehow locate him in 1-5 on the total list of possibilities:

- 1. A does not act.
- 2. A acts, but not voluntarily.9
- 3. A acts voluntarily, but does not cause B's death.
- 4. A's act is justified homicide.
- 5. A's act is excused homicide. 10
- 6. A's act is criminal homicide:
 - a. involuntary manslaughter.
 - b. voluntary manslaughter.
 - c. second-degree murder.

Several possible innovations in legal doctrine are discussed in section IV. The most promising line suggested by standard accounts of

 9 When A ducks to avoid a peril caused by X, as in the squib case, A's act is often called nonvoluntary, in the sense of either "instinctive" or "compelled." See Hart and Honore, op. cit., pp. 146-162, 329-340. This description is used to explain why X is still a cause of the harm to B; hence it might employ different criteria from a voluntariness test for A's own liability. In any case, we ignore "instinctive" ducks (and sacrifices) because the puzzle arises equally for deliberate ones. We assume that nonvoluntariness in the sense of "compulsion" amounts to the defense of necessity or duress.

¹⁰ Criminal defenses are traditionally divided into justifications and excuses, although some defenses are hard to classify. Necessity and public duty are examples of justifications; duress and mistake of fact are excuses. The justification/excuse distinction sometimes has practical effects: a justified, but not an excused, act may be assisted or repeated by other agents; an excused, but not a justified, act of violence may be properly resisted. See, for example, Paul H. Robinson, Criminal Law

Defenses (St. Paul, MN: West, 1984), vol. 1, p. 165.

criminal law is as follows. One discards the artificial distinction between duress and necessity; one also adopts a modern trend to excuse acts under duress that a "person of reasonable firmness" would find irresistible. A's duck can then be held to be excusable homicide under thus unified and liberalized duress/necessity. This line frees A from criminal guilt, with the residual oddity of leaving him a killer. But it does nothing to distinguish ducking from forced shielding. The point of the "reasonable firmness" test is precisely to excuse killing under threat of being killed, to recognize that people of ordinary firmness will sacrifice others' lives to save their own. A does this equally in duck and shield cases. So the legal puzzle that remains is the same as the moral one: A is still either legally guilty or legally innocent in both acts.

IV.

We now consider six ways of solving the basic puzzle. The first two, fair competition and included wrong, are common-sense ideas with no ties to any particular moral theory. Two more belong to action theory, the basic legal metaphysics of action and causation. The last two, the psychological approach of double effect and Kantian exploitation, use specific moral traditions. A skeptical seventh solution is discussed in part v.

1. Fair competition. A natural (if obscure) proposal is that, by sacrificing B, but not by ducking, A unfairly shifts the odds of the competition. Two campers fleeing a bear run a fair race, but perhaps hobbling one's tentmate is as unfair in life as in sport. Everyone in the theater line has an equal chance to notice the gunman and escape, unless shoved directly into the bullet's path. The underlying idea here may be that, if B falls to bullet or bear, B's fate is his fault rather than A's. Fairness explains the profit factor (v): the draft dodger shares the benefits of war, but unjustly avoids the cost of the struggle.

The problem with this proposal is that often B has no chance of escape at all; therefore he has no fair chance; therefore A's using him as a shield cannot deprive him of a fair chance. If A is a champion sprinter, B never had a chance to win the race. Nor did B ever have a fair chance to become a champion sprinter, since athletic ability is partly genetic. To say that B may escape by superior intellect only replicates the problem on the intellectual level. In queue examples, order in line and timing of threat are random, but only the last person in line may be able to escape. A may be taller than B and block his view; only the last car in line may be maneuverable. In these cases the very situation, so to speak, is unfair to B. It is hard to see,

however, how A's act increases the unfairness of a competition that B is certain to lose anyway.¹¹

2. Included wrong (innocence factor). Another idea is that factor (ii) solves the whole puzzle. Tying B's legs and shoving him are wrongs per se—moral offenses under normal circumstances, and crimes of battery. Running or leaping is an act innocent in itself. Perhaps one may escape from harm at B's expense if and only if one does not commit an intrinsically wrong act in doing so.

This proposal has several serious independent defects. First, hobbling and shoving are trivial wrongs and, as batteries, mere misdemeanors. It is implausible to hold that the difference between no wrong and a trivial wrong escalates, when each leads to death, into the moral difference between justified act and criminal homicide. No such multiplier effect is recognized in other contexts. It is no worse —surely not significantly worse—for Judas to betray Jesus with a pinch than with a kiss. It is no worse to frighten B over a cliff by screaming a racial epithet than " $E = MC^2$." It is no worse to feed one's diabetic wife cyanide than placebo insulin.

It is true that traditional law contains at least two examples of multiplier effects, but each is a perennial target of criticism. The first example is the felony-murder and misdemeanor-manslaughter rules. The felony-murder rule at common law—now abolished in England, but surviving in almost all American states—held that a person who causes an unintended death in the commission of a felony is guilty of murder. A similar rule made death caused in a misdemeanor or other unlawful act manslaughter. The moral objection to these rules was well stated by Oliver Wendell Holmes when he suggested that shooting at another's chickens is no more blameworthy when the bullet accidentally hits a person. Holmes wrote:

If the object of the rule is to prevent such accidents, it should make accidental killing with firearms murder, not accidental killing in the effort to steal; while, if its object is to prevent stealing, it would do better to hang one thief in every thousand by lot.¹³

The Common Law (Boston: Little & Brown, 1881), p. 58.

¹¹ One may object that this criticism proves too much: that no one ever loses a fair race. But, even in sport, contests are only considered fair to the degree that there is some possibility any contestant can win. A heavyweight vs. bantamweight wrestling match is not a fair bout. So, if A is a track Olympian and B an arthritic one-legged cripple, even by sport standards the race is unfair, and A scarcely makes it more unfair by hobbling B more. In line examples, the race analogy fails anyway.

¹² LaFave and Scott, op. cit., pp. 545–561, 594–602. The felony-murder rule is often presented as a theory of an "implied" intent to kill, but this is an unconvincing legal fiction. It seems better to regard the rule as a frank multiplier effect.

After heavy criticism by theorists, the rules were abolished in the English Homicide Act of 1957 and the American Model Penal Code.

A second multiplier effect is the traditional "eggshell skull" rule that criminals and tortfeasors "take their victims as they find them."14 This is the doctrine that one is the legal cause of death resulting from one's light blow to a hemophiliac or other unforeseeably fragile victim. In American criminal law, the eggshell-skull rule is normally combined with felony murder or misdemeanor manslaughter; it supplies causation while they supply mens rea. The tort counterpart holds one liable for all damage resulting from negligent conduct acting on existing (but not subsequent) conditions, however bizarre. These doctrines find liability for unforeseeable harm and thus magnify small faults into large ones. To view abnormal subsequent events as excuses, but not abnormal existing conditions, seems irrational, and the eggshell-skull doctrine is not explicitly preserved in the Model Penal Code. Since duck and sacrifice deaths are foreseen, the relevance of these legal multiplier rules is dubious in any case.

Apart from multiplier effects, the included-wrong view of ducking makes a distinctly false prediction: that, when ducking involves a greater intrinsic harm than forced shielding, ducking is worse. Suppose that A's flight must damage C's valuable flowerbeds; it surely does not follow that A should instead shove B or tie his legs. Finally, our intuitions about forced shielding seem largely unchanged when the included act is innocuous. Suppose A uses no force on B, but summons him into a bullet's path by a greeting or a handshake. Or suppose A forcibly moves B not by a shove, but by a friendly arm around his shoulders. Should one then say that A is guiltless, and also guiltless if B ordinarily does not mind being shoved? These examples suggest that to seize on included wrongs is an act of theoretical desperation in the attempt to explain firm but mysterious intuitions.

3. Act versus omission. Some philosophers and jurists will wish to subsume the sacrifice/duck distinction under the act/omission dis-

an act of betrayal. But the betrayal must be either A's insincerity or the death of B. Handshakes are so conventional that "insincerity" is hardly a wrong; and to appeal to B's death makes the explanation circular and destroys the contrast with ducking.

¹⁴ Hart and Honore, op. cit., pp. 172–176, 342–347; Hall, op. cit., pp. 254, 257–261. William L. Prosser, Handbook of the Law of Torts (St. Paul, MN: West, 4th ed., 1971), pp. 261/2. When death is foreseeable, any degree of homicide can be committed by an act that is per se lawful (lawful in circumstances where it does not unjustifiably cause death). One can murder by lying to a blind man approaching a cliff, opening a window in a sickroom, throwing a main power switch, etc.

tinction, or related ones such as doing/allowing.¹⁶ If ducking is an omission, the desired contrast follows from standard legal principles:

Generally one has no legal duty to aid another person in peril, even when that aid can be rendered without danger or inconvenience to himself. He need not shout a warning to a blind man headed for a precipice or to an absent-minded one walking into a gunpowder room with a lighted candle in hand. He need not pull a neighbor's baby out of a pool of water or rescue an unconscious person stretched across the railroad tracks, though the baby is drowning or the whistle of an approaching train is heard in the distance (LaFave and Scott, 183).

Admittedly it is odd to describe running for dear life, or violent leaping, as nonacts. But the suggestion is not absurd, because typical omissive conduct is also an act, the act of doing something else. ¹⁷ In the Samaritan tale, priest and Levite do something, viz., walk on; the pool lizard sips his Margarita while the baby drowns. Conversely, on one classical legal conception of an act as willed muscular contraction, standing still is an act: it requires many muscles and, in the face of a bear or bullet, a will of steel. ¹⁸ Thus, one might view running or leaping as an omission, the failure to stand still. Certainly ethicists and judges often prove flexible in adjusting their act/omission distinctions to yield favored results. ¹⁹

¹⁶ The doing/allowing distinction (e.g., killing/allowing to die) is discussed by Philippa Foot in "The Problem of Abortion and the Doctrine of the Double Effect," Oxford Review, v (1967). Foot suggests that allowing is not the same as omission, since one can allow by acting (removing a plug allows water to flow) and do by omitting (a missing actor spoils a performance). As shown by Nancy Davis, exactly what distinction Foot thinks has moral force is unclear, since she seems to hold that allowing one's child to starve is a kind of murder ["The Priority of Avoiding Harm," in Bonnie Steinbock, ed., Killing and Letting Die (Englewood Cliffs, NJ: Prentice-Hall, 1980), pp. 172–214].

Because omissive conduct is almost always an act or acts, the basic concept is not omission, but causing result R by omission (during time period P). A simple test that usually works well is the following: A causes R by omission if and only if (i) A could have acted in such a way that R would not have occurred, but did not; and (ii) R would also have occurred if A had done no action at all. (Doing nothing is not behaving, e.g., sleeping or falling unconscious.) This test is too narrow when A's inaction would cause others to act (e.g., they notice A's fall).

¹⁸ See Holmes, op. cit., p. 91: "An act is always a voluntary muscular contraction, and nothing else." Although Holmes is in the tradition of John Austin, Lectures on Jurisprudence (London: Murray, 1863), Austin and other writers usually speak of willed bodily movements, which would not cover standing still. It being agreed that there are no purely mental crimes, the issue usually discussed is whether "act" includes effects and circumstances (killing by firing a gun) or only bodily movement

(crooking the forefinger). Standing still as an act is left undiscussed.

19 Foot's construal of starving one's child as murder by allowing has already been mentioned. In law, Fletcher cites Rex v. Smith, 2 Car. and P. 448, 172 E.R. 203

An especially tempting precedent may be the law's treatment of "pulling the plug" on terminal patients as an omission. In an influential article, ²⁰ George P. Fletcher proposed to view disconnecting a respirator as ceasing treatment, and ceasing treatment as nontreatment. If pulling the plug is an omission, it is nonculpable if consistent with the special duty of care imposed by the doctor-patient relationship, which in turn might depend on common practice. One difficulty with Fletcher's position is that a bystander's or intruder's disconnecting the respirator is murder, so that the same conduct is act or omission depending on who engages in it. Another difficulty is that the law normally views ceasing aid more harshly than failure to begin aid.²¹ The former difficulty would not affect the omission theory of ducking, but the latter might.

A more serious problem is that ducking is clearly intermediate between act and true omission. In the gunman case, the true omission is failure of A, standing outside the line, to run *into* the bullet's path, i.e., to make himself a shield. Thus, the omission theory recognizes two omissions, which may be problematic if one is worse than the other—a possibility illustrated below. One might reply that it is no worse to recognize two omissions than two acts of homicide, as ethics and law already do. In the gunman case there are four possibilities:

- 1. A shoots B himself.
- 2. A shoves B into the bullet's path.
- 3. A ducks the bullet to B.
- 4. A fails to run into the bullet's path to shield B.

Perhaps this is a list of two acts (1, 2) and two omissions (3, 4) in order of decreasing culpability. There are other possibilities: A may duck under B's legs (2.5), using B as a shield without laying violent

²⁰ George P. Fletcher, "Prolonging Life: Some Legal Considerations," Washington Law Review, XLII (1967): 999-1016.

The first difficulty mentioned in the text is cited by Steinbock in her introduction to Killing and Letting Die (op. cit., p. 6).

⁽Gloucester Assizes, 1826), involving the three Smiths' injurious confinement of their idiot brother. The court won its way to an acquittal by finding both that keeping the brother locked up was an omission and that the Smiths had no duty of aid. The case antedates, however, modern rules on omissive crime.

²¹ Although one "may have no duty to pick an unconscious person off the railroad tracks as a train is approaching around the bend, yet if he once lifts him off the track, he cannot thereafter lay him down again in his original position; if he should do so, and the train should kill him, he would be criminally liable for the homicide" (LaFave and Scott, op. cit., p. 185). See also Prosser, op. cit., pp. 343/4.

hands upon him. These and other cases²² may suggest a continuum between pure act and pure omission, with various forms of shields, ducks, and duck-shield combinations in between, and moral culpability varying continuously. A continuum of guilt is a truism in ethics, and, although legal offenses are discrete, there is a continuum in punishment.

Unfortunately for this lovely theory, some ducks clearly are culpable homicides. Suppose A's car is parked between a rolling rock and a helpless child. A sees the child in the rock's path, but drives off to avoid an ugly dent. A can be convicted of murder or manslaughter in any American court. On the other hand, had A merely failed to interpose his car between rock and child, he would be innocent in both Anglo-American law and, presumably, an ethics that takes the act/omission distinction seriously. As regards action and causation, in the first case A kills the child by the familiar method of unblocking a hazard. One can kill a man by tampering with his parachute, stripping insulation from his power tool, puncturing his sterile suit in a biological warfare lab, or blasting his submarine hull. A, by moving his car, enables the rock to hit the child, and so, in principle, he also does so when the block is A's own body or body part.

The existence of duck homicides is a fatal objection to the omission view of ducking reminiscent of the first problem with Fletcher's

²² Another possibility occurred in an Oklahoma post office massacre on August 20, 1986. One employee fleeing the gunman locked himself in a spacious vault and survived, but others who were locked out died. This is a case of hogging a refuge. Hogs are worse than ducks, even if one is unbearably cowed.

The fact that the car is A's property is no excuse; one cannot use one's property to kill. Suppose A and B are mountaineers, and B borrows A's pitons without permission. Clearly A cannot repossess them while B is hanging from the rockface. As mentioned above, an otherwise lawful act that foreseeably causes death without

justification or excuse is criminal homicide.

Besides such rare situations, to deny that moving a car aside is an act causing death allows the following ready murder method. Induce *B*, your enemy, to drive behind your van in the left lane of a divided highway with occasional crossroads, such as route 1 south of Media, PA. When a car ahead stops for a left turn, delay changing lanes until the last second. *B*, whose view is blocked by your van, will crash and die.

²⁴ Of course, one can assign moral significance to act/omission in some cases, but not in all cases. Richard Trammell discusses the possibility that acts are worse than omissions only when the omitted act requires more than minimal effort. See "Saving Life and Taking Life," this JOURNAL, LXXII, 5 (March 13, 1975): 131–137. Fixing a dent requires more than minimal effort and expense; what guides intuition in the rock-child case is not the absolute cost to A, but the relative value of car and child.

This is Foot's term for a type of allowing that is removing an "obstacle which is ... holding back a train of events" (Foot, op. cit., reprinted in Steinbock, op. cit., p. 161).

Digitized by Arya Samaj Foundation Chennai and eGangotri

view. The same conduct must be act or omission depending on how much harm it avoids. When A ducks to save his life, it is to be an innocent omission; when A ducks to save his car, it is culpable homicide. But how can a life-saving duck be less an action than a car-saving duck? Quantity of harm ill-suits the basic legal metaphysics of action; surely it belongs rather to the moral calculus of criminal defenses.

4. Causation. Another approach via legal metaphysics appeals to causation: to deny that ducking causes death. Attacking the standard necessary-condition view of causation in law, Fletcher pursues this new line on omission cases in a more recent discussion. ²⁶ It is independent of act/omission in that one may cause death by omission (starving a child). Instead, Fletcher tentatively correlates cause/no cause with killing/letting die, the distinction he likes best for law. Common sense is indeed less inclined to say A kills B by ducking than by forcibly making him a shield. One feels that to grab B is to do something to B, while in ducking A does not act on B at all. Fletcher observes that similar concepts of interference or intrusion are basic to the ethical frameworks of John Rawls and Robert Nozick. ²⁷

How to define this "relational theory of action" (592) is one major challenge to Fletcher's view. In grabbing or hobbling, A makes physical contact with B, but A may also summon B by greeting or gesturing without contact. Greeting B pushes sound waves upon him, but so does leaping aside. No doubt the sound waves of a greeting are part of a direct causal chain leading to B's death, whereas the sound waves of A's leap are not even physically necessary conditions for it. Yet in (c2) Alison acted only on Babette's briefcase, not directly on Babette. Our cases suggest that the concept of acting on, even if ultimately physical, is not easy to analyze.

Another grave problem is the same objection made to the act/omission theory. It is unclear how the unblocked-hazard kind of killing fits Fletcher's approach. If one can kill by unblocking another force and if killing requires acting on, then unblocking a force that

²⁶ Rethinking Criminal Law (Boston: Little, Brown, 1978), esp. pp. 588–610. ²⁷ Ibid., pp. 603/4. According to Fletcher, Rawls requires a concept of interference to explain the first principle of justice and to distinguish it from the second. One may restrict conduct that interferes with the liberty of others (e.g., religious persecution but not religious liberty); but nonallocation of resources to a person is not an interference, hence not a restriction of his liberty [Rawls, A Theory of Justice (Cambridge, MA: Harvard, 1971), pp. 60, 195–257]. Nozick uses "impinge" in discussing boundary crossings [Anarchy, State, and Utopia (New York: Basic Books, 1974), pp. 57–87].

acts on *B* is acting on *B*. Thus, in the same cases, whether a duck causes death must depend on the amount of harm *A* avoids. To solve the puzzle, the causal theory must say that a car-saving duck causes *B*'s death, but a life-saving duck does not, which is as implausible as calling only the former an act. Again, such quantitative considerations belong in criminal defense, not in basic law of act and causation.

- 5. Double effect (psychological solution). A difference in A's intent in duck versus sacrifice could solve the puzzle for both ethics and law. Perhaps, in sacrificing A intends B's death, whereas in ducking A foresees it but does not intend it. The most famous doctrine of this sort is the doctrine of double effect, used in Thomistic ethics but not in Western law. In current form this doctrine states that an act with both good and bad effects is permitted if four conditions hold:
 - (i) The act is not wrong in itself.
 - (ii) The good effects are proportionate to the bad.
 - (iii) Only the good effects are intended.
 - (iv) The bad effects are not the means to the good.28

Condition (iv) is a gloss on (iii), saying that whoever intends the end intends the means. In some formulations, condition (iii) says that only the good effects are "desired."

In some ways double effect is more successful with ducking than other views. It handles the gunman case, since, if B is a human shield, B's death is A's means (iv). When A leaps aside, B's death cannot be the means to A's survival, since leaping saves A's life even if no B is behind him. This does not hold in the bear story, however, since the bear is faster than both A and B, and therefore A survives only by B's death. So one would expect Thomistic ethics to approve ducking bullets but not bears, 29 except that Catholic applications of double

²⁸ See, for example, "The Principle of Double Effect," The New Catholic Encyclopedia (1967); Germain Grisez, Abortion: The Myths, the Realities, and the Arguments (New York: Corpus Books, 1970). Some major disagreements about double effect occur among contemporary Catholic writers, especially over the means condition. On the concept of means, see Grisez, p. 330 ff., and Nancy Davis, "The Doctrine of Double Effect: Problems of Interpretation," Pacific Philosophical Quarterly, LXV, 2 (April 1984): 107–123.

²⁹ Grisez's new theory (op. cit.) of double effect loosens the orthodox tests for unintended side effects, accommodating Aquinas's original thesis (Summa Theologiae, Treatise on Justice, question 64, article 7) that self-defense is unintended killing (not justified intentional killing). Grisez holds that any foreseen effect can be unintended if it is physically inevitable, i.e., "humanly indivisible" from the good effect. Grisez thus applies Aquinas's self-defense view to all cases, even suction

effect are often capricious. Double effect also easily condemns a fatal car-saving duck as violating proportionality (ii): a property value is not proportionate to a child's life. On the other hand, in scarce-resource cases, to "sacrifice" *B* is not to enshield him, and therefore the means concept may fail to explain the distinction.

In other cases double effect definitely diverges from intuition, simply yielding a strange classification. A's total plan in ducking a bullet can include B's death, if A's wife C is behind B. In this situation A must desire B's death (or injury) as a means to his wife's survival, which violates at least the desire version of (iii) and perhaps (iv). Nevertheless it is hard to feel very indignant at such a duck. Conversely, as Charles Fried observes, double effect may justify shoving people to their deaths in one's headlong flight. Since one would have acted the same if they were not in the way, one passes a standard Catholic test of intent. Clearest and most counterintuitive is the case of ducking under B's legs. Here B's death or injury is again A's means of survival, violating (iv). Yet it seems preposterous to claim that leaping sideways is innocent but leaping over or under B is murder.

In three-person cases it is debatable whether double effect's judgments are correct. When Y shoves A out of danger that strikes B, Y's motive is crucial. If Y's intention is to save A, the act is permissible; if his intention is to harm B, the act is wrong. In the latter case, pushing A aside is just as bad as pushing B in front—i.e., the duck/sacrifice distinction disappears. In general, double effect approves benevolent and condemns malevolent victim selection. Perhaps that is right. Or perhaps an alternative consequentialist view is better: that, if the act is right, bad motives cannot make it wrong. On that view, in some happy situations one may exercise evil impulses without moral cost. Such is the law's view, insofar as one who knows he has a valid defense may kill out of hatred with impunity. One may lawfully delight in killing one's archenemy in self-defense or by judicial execution:

Whenever the circumstances of the killing would not amount to murder, the proof even of express malice will not make it so. One may harbor the most intense hatred toward another; he may court an opportunity to take his life; may rejoice while he is imbruing his hands in his heart's

abortion, relaxing a few traditional prohibitions and redescribing many cases of evil means as failures of proportionality. In particular, he holds (334) that, without violating the ban on suicide, a mother may "save her child by purposely interposing her body as a shield against an attacking animal." So Grisez's version of double effect probably approves ducking bears as well as bullets.

blood; and yet, if, to save his own life, . . . he was fully justified in slaying his adversary, his malice shall not be taken into the account. The principle is too plain to need amplification. *Golden v. State*, 25 Ga. 527,532 (1858).

This view simply illustrates the law's general tendency to reject motive in favor of other factors, e.g., intent. Since Catholic ethics now admits the intent to kill in these cases, however, such cases are not strictly analogous to the double-effect view of ducking.

6. Kant. A sixth solution, which overlaps the Catholic approach, is to employ the Kantian concept of "using a person as a means." Kant's second version of the categorical imperative holds that rational beings must always be treated as ends, never merely as means. As in double effect, if B is a shield, his body is the means to A's survival. Kant would say that this act denies B's humanity, that it uses or exploits him. Nancy Davis suggests this explanation for many of our intuitions about act/omission cases.

Like double effect, a Kantian view of exploitation condemns ducking under B's legs. It may also have the extra drawback of condemning ducking behind B in a two-shooter case. Suppose two gunmen are each about to fire at A and B respectively. It seems obvious that (at least if A can and B cannot) A may rightly save his life by ducking behind B, who will die anyway. In thriller argot, B is dead on his feet. But this action treats B's body as a means to survival. Perhaps one can reply that here B's body is a means, but not B's death, which will occur anyway, and therefore not B himself. This reply may excuse a two-shooter duck. In the one-shooter case, however, if the choice is between ducking and sacrificing, then B's death is also inevitable relative to this choice—B is relatively dead on his feet. Therefore, one ought to be allowed to exploit B's body, as in the two-shooter case; therefore, sacrificing should be as moral as ducking. Insofar as this argument succeeds, Kantians avoid the wrong two-shooter result at the cost of destroying the duck/sacrifice distinction in the one-shooter case.

Another objection to the Kantian solution is its restriction to harm to persons. One is forbidden to exploit persons but not objects. On the Kantian view, the duck/sacrifice distinction should disappear when the harm is to an object—or to a nonrational being, like an animal. This seems to violate intuition, except that all intuitions weaken as values shrink. Smith may move his car from the path of a rockslide, but may not shield it with Jones's car. Smith may spray his beans with Japanese-beetle repellent, but may not spray Jones's beans with Japanese-beetle attractant, even if the same beetles end up on the same beans. Kantians will say that to invade a person's

property is to exploit the person. This move is necessary anyway to condemn theft; so Kantian ethics is no worse off for these examples. But does the distinction vanish with unowned objects, such as wild animals? Is there a difference between pulling one's dog away from a bear, diverting his appetite to a wild dog, and throwing the wild dog to the bear? If there is, a Kantian view of exploitation fails to solve the puzzle. The problem equally plagues a Kantian approach to act/omission, where it is clear that our intuitions survive a switch to animals.

V.

A final solution is the skeptical one: to deny that the duck/sacrifice distinction *per se* has any moral significance. At least four arguments might attack the distinction. First is the failure of all six theories to match intuition. If no theory handles all ducks and sacrifices, perhaps our intuitions are irrational. Such a judgment might be reasonable after a long history of failed analyses, but is now, to say the least, premature.

A second reason for skepticism is our thesis that ducking is not always innocent. If some ducks are culpable homicides, but corresponding pure omissions are not, then one might argue that ducking is not per se significant in the way omission is. But that is to argue that, because our distinction is not always important, it is never important—a non sequitur. Defenders of act/omission also find their distinction sometimes unimportant. It is true that the culpable (nonstatutory) omissions of traditional law involve special relationships, not degree of harm as in the rock-child case. But European Good-Samaritan laws and some ethicists (e.g., Richard Trammell) do classify omissions by degree of harm. Furthermore, even for culpable ducks, the distinction may be important. Many people would judge more harshly a man who guards his car by throwing a child into a rock's path than one who only drives his car away, exposing the child. Thus, the duck/sacrifice distinction may remain as a matter of degree.

A third argument for skepticism is the duck/shield continuum imagined in section IV. Somewhat artificially, a continuum can be constructed, and one of a type that inspires a sorites. One way to do so in the bear story is for Alex to pour water on Bruce's boots, the camp being in rugged terrain. If Alex soaks Bruce's boots so thoroughly that Bruce cannot run at all, his act is pure sacrifice. If he leaves the boots alone and runs, it is pure duck. In between he can drip one drop on the boots, or two drops, or three drops, and so on.

Still, a similar continuum can be constructed between pure act and pure omission. Suppose *B* bleeds to death from two wounds, *A*'s stab

and a tiger bite, and, as *B* bleeds to death, *A* refrains from carrying him to a doctor. If *A*'s stab is so violent that *B* bleeds too fast to be saved, *A* kills him by pure act. If *A* does not stab *B* at all, *A*'s conduct is pure omission. In between, *A*'s stab wound can bleed one drop, or two drops, and so on. The conditions for a sorites are met here by the legal maxim *de minimis non curat lex*, which implies that accelerating *B*'s death by one drop is not killing, but at some indeterminate point *A* becomes a killer. Another example is the way Gulag prisoners die from combined starvation and overwork. The fact is that many distinctions, like day/night, survive a continuum; to think otherwise is an instance of the twilight fallacy. A continuum no more refutes the duck/sacrifice distinction than it refutes act/omission or any other ordinary distinction; it merely shows them to be vague, like all distinctions in ethics and law.

The fourth argument for rejection would be to supplement the first by explaining our intuitions away. This could be done in either of two ways. One might ascribe them to irrational features of human moral psychology—e.g., the superstition about identifiable victims. Or one might explain them as due to other genuine factors with which the duck/sacrifice distinction usually covaries. The latter is the main line of writers who deny significance to the act/omission distinction. They hold that cases where act/omission seems important really hang on factors such as motive, intent, certainty of result, ease of avoidance, and number of omitters.

Of these factors, certainty, intent, and motive are relevant to ducking. As to certainty, one stipulates hypothetically that Alex is sure to outrun Bruce, but such certainty quickly evaporates from the mind. In real life Alex is not sure; he may not even think the problem through, merely running as hard as he would if camping alone, but, thanks to Bruce, with greater success. Panic and uncertainty negate intention. If we produce certainty by making Bruce a helpless quadriplegic, common sense wavers. Now hobbling Bruce is no worse than outrunning him, and it may be heartless of Alex not to knock Bruce unconscious before running away, to spare his agony. As to intent and motive, whenever A chooses sacrificing over ducking, A's intent to kill is clear and his motive suspect. But, if A must sacrifice B or die—e.g., the gunman scenario in a narrow corridor—the wrongness of sacrificing is less obvious.

³¹ Perkins and Boyce, op. cit., pp. 779/80. Hart and Honore (op. cit., pp. 351/2) express some uncertainty about this application of de minimis.

^{\$2} For example, Michael Tooley, "An Irrelevant Consideration: Killing vs. Letting Die," pp. 56–62 of Steinbock, ed., op. cit.; Peter Singer, Practical Ethics (New York: Cambridge, 1979).

Nevertheless we are unready to embrace the skeptical solution. Unless one is a doctrinaire utilitarian, intuitions so firm and prevalent should not be surrendered without a struggle. What we claim is that the duck/sacrifice distinction is a package deal with act/omission—an act/omission analogue within the realm of acts. As with act/omission, moreover, to reject a distinction in ethics may not be to reject it for law. In general, criminal law should approximate social moral consensus, if only to avoid capricious enforcement and the chaos of jury nullifications of law in mercy killing and insanity cases. Other nonmoral policy considerations supporting act/omission may support duck/sacrifice as well. Thus, it may be that criminal law must treat ducks more kindly than sacrifices even if ethical theorists disagree.

Our aim in this paper has been to state a puzzle and survey possible solutions. We have focused throughout on two-person or threeperson cases, leaving multiparty cases and social policy aside. It is worth noting that ducking and sacrificing occur in the national and international arenas, especially in war. The history of World War II is one long series of ducks and sacrifices by allies and enemies.³³ In domestic policy, distribution of harms and benefits-which, unlike our simple examples, come in countless types and magnitudes often involves choices between direct and indirect injuries to various groups. The issue whether government penalties on a constitutionally protected choice (abortion) are equivalent to rewards for choosing the alternative (maternity care funds) might be seen as a duck/ sacrifice choice by a third party, government.34 Many defenders of affirmative-action quotas seem convinced that racially reserved places, as in Bakke, directly benefit minorities but only indirectly harm whites. We invite other writers to explore social applications of the distinction.

CHRISTOPHER BOORSE

University of Delaware

ROY A. SORENSEN

New York University

33 As an example of a duck, the United States was criticized for adopting strategies that maximized Russian casualties. An example of a sacrifice is the radio communication deliberately initiated with a shipping convoy near Gibraltar in fall 1942 to divert German submarines from attacking Patton's North African Torch invasion fleet. The examples are interesting in that usual moral judgments are reversed: the duck is condemned, the sacrifice approved.

⁵⁴ This was the issue in the abortion funding cases (Harris v. McRae, 448 U.S. 297, 1980; Maher v. Roe, 432 U.S. 464, 1977). Comparing abortion to private schools, the Supreme Court ruled that government cannot prohibit abortion, but

need not fund it equally with maternity.

TACHYONS, TIME TRAVEL, AND DIVINE OMNISCIENCE

OR philosophers in either field, philosophy of science and philosophy of religion are too often viewed as mutually irrelevant disciplines. As a result, insights acquired in each field may not be appropriated by philosophers working in the other field. This is unfortunate, because sometimes the problems can be quite parallel and a consistent resolution is required. One especially intriguing case in point concerns, in philosophy of science, the possibility of tachyons and time travel and, in philosophy of religion, the relationship between divine foreknowledge and human freedom. It is rarely appreciated by discussants of these respective issues that the problems are quite parallel and that insights garnered in the resolution of the difficulty in one discipline may have provocative implications for the solution of the parallel problem in the other field.

I. THEOLOGICAL FATALISM

To begin, then, with philosophy of religion: Greek fatalism, embodied in Aristotle's argument of De interpretatione 9, posed a special threat to Christian theology. Committed to the biblical doctrine of divine foreknowledge as well as to human freedom, Christian thinkers had to explain how it is either that God knows future contingents without future contingent propositions' being antecedently true or false or that God's knowing the truth value of such propositions does not after all entail fatalism. The problem of theological fatalism seemed especially acute since God's foreknowledge of some future event is itself a fact of past history and therefore temporally necessary; that is to say, it no longer has any potential to be otherwise. Therefore, what God foreknew must necessarily come to pass, since it is impossible that God's knowledge be mistaken. In our own day, philosophers such as A. N. Prior, Richard Taylor, Steven Cahn, Nelson Pike, and Paul Helm have argued that from the temporal necessity of

1. God foreknew p.

and the logical necessity of

2. If God foreknows p, then p.

it follows, for any future-tense proposition p, that necessarily p.

The majority of contemporary philosophers have, however, disputed the cogency of such reasoning. From the fact that God foreknows that I shall do x, it follows, not that I cannot do otherwise, but only that I shall not do otherwise. It remains within my power not to

0022-362X/88/8503/0135\$01.60

© 1988 The Journal of Philosophy, Inc.

do x, but, given God's foreknowledge, we know that I shall not in fact exercise that power. Were I to do otherwise, then God would have known different future-tense propositions than He in fact knows. As for so-called "temporal necessity," this notion is notoriously difficult, and, if this is a legitimate kind of modality, it is not at all evident that God's foreknowledge of some future event is characterized by such necessity. This does not mean that it is within one's power to change the past. Rather it is to assert the truth of the counterfactuals

- 3. If I were to do x, God would have foreknown that I would do x. and
 - 4. If I were not to do x, God would have foreknown that I would not do x.

From the fact that God foreknows that I shall do x, we may therefore infallibly infer that I shall do x, but it would be fallacious to infer that it is not within my power to refrain from doing x.

II. TACHYONS

This rejoinder to theological fatalism, which seems to me altogether correct, has some disturbing consequences when we turn to philosophy of science to investigate the possibility of tachyons and of time travel. When Albert Einstein proposed his Special Theory of Relativity in 1905, he conceived of the speed of light c as a limiting velocity such that transmission of energy from point to point in space-time at superluminal velocities is impossible: "velocities greater than that of light," he concludes, "have no possibility of existence." This is because the mass of a particle would become infinitely large as its velocity approaches c. The speed of light was therefore conceived to be an inviolable barrier for particle velocities. In the second half of

³ 'Elektrodynamik bewegter Körper,' Annalen der Physik, xvII (1905):

¹ For the clearest statement of this position, see Alvin Plantinga, God, Freedom, and Evil (New York: Harper & Row, 1974), pp. 69–72; for an assessment of this solution, see Philip Quinn, "Plantinga on Foreknowledge and Freedom," in James Tomberlin and Peter Van Inwagen, ed., Alvin Plantinga, Profiles 5 (Boston: Reidel, 1985), pp. 271–287.

² For the best analysis, see Alfred J. Freddoso, "Accidental Necessity and Logical Determinism," this JOURNAL, LXXX, 5 (May 1983): 257–278; cf. Joshua Hoffman and Gary Rosenkranz, "Hard and Soft Facts," *Philosophical Review*, XC (1984): 419–434; Alvin Plantinga, "Ockham's Way Out," *Faith and Philosophy*, III (1986): 235–269. On none of these theories of temporal necessity does God's foreknowledge turn out to be temporally necessary. For an assessment of these theories, see my "Temporal Necessity; Hard Facts/Soft Facts," *International Journal for Philosophy of Religion*, xx (1986): 65–91.

the century, however, physicists such as Olexa-Myron Bilaniuk, V. K. Deshpande, E. C. George Sudarshan, and Gerald Feinberg realized that Einstein's conclusion was overdrawn. Although his equations prohibited the *acceleration* of particles traveling at subluminal velocities to or beyond c, they did not preclude the existence of particles whose velocities are *always* greater than or equal to c. After all, photons and neutrinos both travel with a velocity equal to c without ever having been accelerated from a subluminal speed to luminal velocity. So why could there not exist particles that travel at superluminal velocities without ever having been accelerated from speeds less than or equal to c? In this case the speed of light remains an inviolable barrier, but that does not preclude the existence of particles on the other side of the barrier. Feinberg dubbed such particles tachyons, from $\tau \alpha \chi \iota s$ (swift), and the experimental search for these exotic entities was on.

And, indeed, if tachyons do exist, they are exotic. Apart from other oddities, the equations for energy and momentum for such particles reveal that tachyons would accelerate as they *lose* energy. Conversely, whenever energy was imparted to a tachyon, it would *decelerate*. This leads to one of the most peculiar characteristics of tachyons: their prima facie possession of negative energy. Let an observer at rest in a reference frame S observe a tachyon traveling with a velocity v relative to him. This same particle will travel with a different velocity v relative to another observer in a reference frame v0 which is moving with respect to v0 with a velocity v0. When the product v0 exceeds v0, the tachyon will possess negative energy relative to v0. More peculiar still, such particles will seem to travel backward in time. To the observer in v0 the negative-energy particle would appear to be absorbed first and emitted later.

The implications of such behavior were noticed by Richard Tolman as early as 1917 in what has come to be known as Tolman's paradox, namely, that communication with the past is possible. Let an observer O in a reference frame S send out a burst of infinitely fast tachyons at t_1 to an observer O^1 in a reference frame S^1 which is receding from S at the uniform velocity w. The reception of the tachyon signal in S^1 triggers a similar burst of tachyons back to O

⁴ Bilaniuk, Deshpande, Sudarshan, "Meta Relativity," American Journal of Physics, xxx (1962): 718ff; Gerald Feinberg, "Possibility of Faster-than-light Particles," Physical Review, CLIX (1967): 1089–1105.

⁵ The Theory of Relativity of Motion (Berkeley: University of California Press, 1917), pp. 54/5. Actually Tolman's paradox results not only when infinite velocities are involved, but for all velocities greater than c^2/w , where w is the relative velocity of two observers

Digitized by Arya Samaj Foundation Chennai and eGangotri which travel with an infinite velocity relative to S^1 . The relativity equations dictate that the second signal arrives in S at a time t_0 before the burst of tachyons is sent at t_1 . But, since the signal from O^1 to Swas triggered by the signal from O to S^1 , it follows that the effect (O's reception of O^{1} 's signal) precedes the cause (O's sending his signal to O^1) in S, or, in other words, tachyons furnish the mechanism for backward causation.

This implication alone was enough to warrant the rejection of the possibility of tachyons in the minds of many physicists. 6 Proponents of tachyons felt at first constrained to explain away Tolman's paradox with its attendant backward causation by means of a "reinterpretation principle." "It is precisely by putting together the two quizzical concepts of 'negative-energy' particles traveling 'backward in time' that the resolution of the difficulty is found," state Bilaniuk and Sudarshan; "A 'negative-energy' particle that has been absorbed first and emitted later is nothing else but a positive-energy particle emitted first and absorbed later, a perfectly normal situation." By interpreting any negative-energy particle moving backward in time as a positive-energy particle moving forward in time, one may thereby eliminate the occurrence of an effect before its cause. In our previous case, for example, O1 will naturally regard the tachyon beam received from S as actually a signal that he is himself sending to S^1 . O^1 and O will regard these beams as spontaneous emissions from their own tachyon transmitters rather than as receptions from another reference frame.

Now, at face value, the reinterpretation principle sounds merely like the endorsement of what can only be characterized as a fantastic delusion. If O's tachyon signal really does trigger O's transmitter to send a return signal, then it is simply irrelevant whether O or O^1 believes that no backward causation has occurred. Perhaps the best face to put on Bilaniuk and Sudarshan's remarks is to interpret them as claiming that the causal relation is itself relative to reference frames; that is to say, there is no absolute causal directionality in the same way that there is no absolute simultaneity according to Special Relativity. The world-line of the tachyon burst simply exists (tenselessly) between space-time points in S and S^1 , and whether the tachyons are moving from S to S^1 or vice versa is observer-depen-

⁷ Bilaniuk and Sudarshan, "Particles beyond the Light Barrier," Physics Today (May 1969): 47; Gerald Feinberg, op. cit., p. 1091.

⁶ See Bilaniuk et al., "More about Tachyons," Physics Today (December 1969), p. 49; David Bohm, The Special Theory of Relativity (New York: W. A. Benjamin, 1965), p. 158; F. A. E. Pirani, "Noncausal Behavior of Classical Tachyons," Physical Review, D 1 (1970): 3224.

dent, as is also which event is conceived to be the cause and which the effect. Unfortunately, it has been shown that, even on this understanding, backward causation cannot be precluded. More to the point, however, the notion that causal directionality is relative to reference frames seems clearly untenable. In their engaging discussion of a tachyonic antitelephone, Benford, Book, and Newcomb point out that causal directionality is independent of temporal considerations and is therefore not susceptible to arbitrary reinterpretation:

For example, let A be William Shakespeare and B Francis Bacon, and let V^1 [the outgoing tachyonic velocity] be negative. If Shakespeare types out Hamlet on his tachyon transmitter, Bacon receives the transmission at some earlier time. But no amount of reinterpretation will make Bacon the author of Hamlet. It is Shakespeare, not Bacon, who exercises control over the content of the message (265).

Thus, "the direction of *information transfer* is necessarily a relativistic invariant. An author's signature, for example, would always constitute an invariant indication of the source" (*loc. cit.*). The reinterpretation principle is thus seen to be essentially an exercise in self-delusion: causal directionality is invariant across reference frames, and to interpret events as related otherwise than as they are is only self-deception.

In light of these facts, proponents of tachyons began to reassess whether backward causation was after all so objectionable or paradoxical. One writers argued that the problem entailed by permitting tachyonic backward causation is fatalism. Feinberg, for example, called this the "most serious qualitative objection" to tachyons; the transmission of signals into the past of a single observer is in apparent conflict with the natural view that one is free to decide whether or not to carry out an experiment up until the time that one actually

⁸ See Bilaniuk et al., "More about Tachyons," pp. 48–50; G. A. Benford, D. L. Book, and W. A. Newcomb, "The Tachyonic Antitelephone," Physical Review, D 2 (1970): 263–265 [this is the same Newcomb of the famous Newcomb's paradox]; Pirani, op. cit., p. 3224; Paul Fitzgerald, "Tachyons, Backwards Causation, and Freedom," in PSA 1970, Roger C. Buck and Robert S. Cohen, eds. Boston Studies in the Philosophy of Science, VIII (Boston: Reidel, 1971), pp. 421–423; T. Chapman, Time: a Philosophical Analysis, Synthese Library (Boston: Reidel, 1982), pp. 23–25.

⁹ Cf. Fitzgerald, "Tachyons," pp. 421–423.

Roger G. Newton, "Causality Effects of Particles that Travel Faster Than Light," *Physical Review*, CLII (1967): 1274. Interestingly, Newton acknowledges his debt to Michael Scriven on the score of causal directionality and time and appeals to tachyons to show the possibility of precognition experiments. See also Paul L. Csonka, "Advanced Effects in Particle Physics, I," *Physical Review*, CLXXX (1969): 1266–1281; Bilaniuk *et al.*, "More about Tachyons," p. 52.

Digitized by Arya Samaj Foundation Chennai and eGangotri

does so."¹¹ The objection seems to be that one could, for example, call oneself in the past on a tachyonic antitelephone and then, after receiving the call, decide not to place it after all. Our discussion of theological fatalism, however, makes the flaw in the reasoning clear: the fact that one has received a call from oneself entails not that one is not free to refrain from placing the call, but only that one will not in fact refrain from placing it. ¹² If one were to refrain from placing the call, then one would not have received it in the first place. Thus, no fatalistic paradox is generated by the existence of negative-energy tachyons.

But, although objections to tachyons based on fatalism are unimpressive, a more substantive objection appears to arise when one considers cases in which tachyonic backward causation would entail the existence of what Paul Fitzgerald has called a "logically pernicious self-inhibitor" ("Retrocausality," 534/5). Benford, Book, and Newcomb invite us, for example, to envisage a situation in which observers A and B enter into the following agreement: A will send at 3:00 a tachyonic message to reach B at 2:00 if and only if he does not receive a message from B at 1:00. B will send at 2:00 a message to reach A at 1:00 if and only if he receives a message from A at 2:00. Therefore, the exchange of messages takes place if and only if it does not take place. They conclude that "Unless some truly radical solution is found to this paradox, we must conclude that tachyon experiments [such as those being currently carried out] can only yield negative results" (265). John Earman points out that such paradoxes

¹¹ Op. cit., p. 1092. Cf. Chapman, *Time*, p. 23, who asserts that, after receiving a return signal which he will trigger, the observer may decide not to send his signal after all; in this case the standard objection to backward causation applies.

¹² Cf. Fitzgerald, "Tachyons, Backwards Causation, and Freedom," pp. 428–434; and, "On Retrocausality," Philosophia, IV (1974): 543. Suppose, he says, I receive a tachyon message from the future that a man I am about to shoot will be at a banquet two days hence. Is it therefore not within my power to kill him? Not at all, responds Fitzgerald; I have both ability and opportunity to do so, so that I could kill him; but were I to do so, I would not have this reliable message from the future that he is alive. The point is that ignorance is not a necessary condition of an action's being within one's power. Fitzgerald's analysis is flawed, however, when he proceeds to argue that, in the case in which one does not try to perform the action precisely because one believes the tachyon message, then one's freedom is limited by the message from the future. For anything, he claims, which prevents a person's doing what he wants to is a limit on his freedom. Fitzgerald fails to see, however, that in this case what one wants to do is changed by the message; it does not therefore prevent one from doing what one wants to do. It merely changes one's motivation. As Fitzgerald goes on to observe, this can arise without messages from the future at all. Suppose before I pull the trigger someone rushes up and informs me that my intended victim is my beloved, long-lost uncle. Suddenly my motivation is changed, and I no longer want to kill him, but would we say that my informer has limited my freedom in conveying his report to me?

do not depend on human agency, but may be constructed solely with machines. Thus, the reinterpretation principle is irrelevant. "A contradiction is generated by asking whether a certain event occurs; we find that it occurs if and only if it does not occur." Although the tachyon event might be interpreted differently by different observers, this difference is totally irrelevant to the contradictory nature of the conclusion.

Now, it is not the existence of tachyons as such, admits Earman, that entails the possibility of a logically pernicious self-inhibitor; rather it is the whole situation which is impossible, and this includes assumptions concerning the possibility of controlling tachyon beams, of detecting them, and so forth. By giving up one or some of these other assumptions, one may impose consistency conditions on hypothetical cases so that the paradox cannot arise. Thus, Fitzgerald maintains that we must conclude only that tachyons cannot be controlled in all the ways required for the self-inhibitor to function.14 When asked why such machines fail, he responds that it may be either for empirical reasons involving constructibility or controllability or owing to a fortuitous set of accidents each time one tries to experiment. The difficulty with the attempt to impose consistency conditions based on considerations of constructibility and controllability, however, Earman explains, is that we have good reason to believe that such devices are possible. The assertion that such experiments cannot be carried out is, therefore, "brazen," since the experiments involve "only operations which we know to be possible in our world."15 Since such devices as are required for these experiments are apparently nomologically possible, it follows that tachyons are nomologically impossible and therefore do not exist. The threat of fortuitous accidents' preventing such experimentation seems utterly implausible, Fitzgerald himself confesses, for we should then have to posit a lawlike regularity of accidents to prevent the functioning of a machine which should be constructible if tachyons exist ("Tachyons," 428). Hence, the conclusion of the foregoing analysis

¹³ "Implications of Causal Propagation outside the Null Cone," Australasian Journal of Philosophy, L (1972): 254. Thus, the escape route suggested by DeWitt, that information sent into the past is wiped from the observer's memory, is unavailing (Bikaniuk, et al., "Tachyons," p. 50).

ing (Bikaniuk, et al., "Tachyons," p. 50).

14 "Tachyons," p. 427; and "Retrocausality," p. 435.

15 Earman, "Causal Progagation," pp. 234/5. Assuming that the apparatus will work as it is supposed to, a typical experiment will involve the following elements: (1) a tachyon source that can be amplitude modulated, (2) a tachyon detector, (3) a velocity filter giving a monoenergetic beam. Proposed devices for each of these are used in tachyon research. (Benford et al., "Antitelephone," p. 263; cf. Bilanuik and Sudarshan, "Particles," pp. 50/1; et al., "Tachyons," p. 52.)

would seem to be that, given the nomological possibility of tachyon emitters and detectors, one cannot avoid the paradoxes by denying assumptions concerning such devices, but is led instead to denying the possibility of the existence of tachyons. Although this reasoning has, to my knowledge, gone unchallenged in the tachyon literature, there is, within the body of literature on the possibility of time travel, a significant challenge to the modal validity of inferring that tachyons are impossible from the nomological possibility of such devices, a challenge akin to the argument against theological fatalism. Let us therefore turn to that discussion.

III. TIME TRAVEL

Long the darling of science-fiction enthusiasts, time travel has come under serious scrutiny in this century. Scientists and philosophers agree that the sort of time machine envisioned by H. G. Wells in his popular novel is in fact an impossibility. Since Wells's machine was conceived to move only through time but not through space, it would, so to speak, "run into itself" as it traveled both forward and backward in time. 16 Moreover, it seemed to involve the contradiction of traversing, say, one hundred years of time in five minutes of time, since it was sitting in the same place. With the development of relativity theory, however, which posited the traveler's relative motion in space as well as time, time travel re-emerged as a new possibility. In 1949 Kurt Gödel drafted a model universe using Einstein's field equations which was similar to Einstein's in that it was both static and spatially homogeneous, but which differed from Einstein's universe in that Gödel assigned a negative value to the cosmological constant (which Einstein had introduced into the equations to prevent the model universe from expanding) and posited an absolute, cosmic rotation of matter, so that isotropy was precluded.¹⁷ On Gödel's model, it was not possible to define a cosmic time because the local times of observers which are associated with the mean motion of matter cannot be fitted together into one world time. The most incredible feature of this model was that it permitted the existence of closed, timelike loops, so that by making a round trip on a rocket ship in a sufficiently wide curve, it would be possible for some observer to

¹⁶ See Monte Cook, "Tips for Time Travel," in Nicholas D. Smith, ed. *Philosophers Look at Science Fiction* (Chicago: Nelson-Hall, 1982), pp. 47–55. See also Donald C. Williams, "The Myth of Passage," this JOURNAL, XLVIII, 15 (July 19, 1951): 457–472, p. 463.

^{17 &}quot;A Remark about the Relationship between Relativity Theory and Idealistic Philosophy," in Albert Einstein: Philosopher-Scientist, 2 vols., ed. Paul Arthur Schlipp (rep. ed.: New York: Harper, 1959), pp. 557–562. Gödel also announced discovery of expansion models and models with any value for Λ for which there exist no cosmic time because of the presence of cosmic rotation.

travel into any region of the past or future and to return. Although the world-line of every fundamental particle was open, so that no temporal period could recur in the experience of an observer connected with the particle, other closed, timelike lines could exist such that, if P and O are any two points on the world-line of a fundamental particle and P precedes O, then a timelike line exists connecting P and O on which O precedes P. By following these loops an observer could fulfill Wells's dream of time travel.

The question is whether Gödel's model constitutes a mere mathematical curiosity or represents a possible description of the real universe. Unfortunately for time-travel buffs, it seems pretty clear that Gödel's universe fails as an actually descriptive account of the universe, and so time travel is not a possibility for us. That is to say, Gödel's universe, even if nomologically possible, is not physically possible. As G. J. Whitrow observes, the empirical evidence for world isotropy undercuts the postulate of cosmic rotation and furnishes instead evidence for the existence of cosmic time. The microwave background radiation is remarkable precisely for its isotropy, which varies by only about one part in a thousand. "Consequently, we have strong evidence that the universe as a whole is predominantly homogeneous and isotropic and this conclusion . . . is a strong argument for the existence of cosmic time." Since these facts are incompatible with Gödel's model, it follows that time travel, at least along his lines, is physically impossible.

But the issue remains whether time travel is not possible in a broader sense. Here the proponents of time travel have argued persuasively that the stock objections to the possibility of time travel are unsound. For example, Gödel himself was disturbed because he believed that his models make it possible that someone might travel into the past and find a person who would be himself at some earlier period of his life. "Now he could do something to this person which, by his memory, he knows has not happened to him" (561). This objection, however, is once again infected by the fallacious reasoning of fatalism. For from the fact that someone did not do something, it does not follow that he could not have done it. Hence, Gödel was unnecessarily concerned about my doing something to myself which I could not remember: all that follows from his objection is either that I did not perform the action or that I forgot it. 19

But at this point a more formidable objection to time travel may be

¹⁸ The Natural Philosophy of Time, 2nd ed. (New York: Oxford, 1980), p. 307. See Paul Horwich, "On Some Alleged Paradoxes of Time Travel," this JOURNAL, LXXII, 14 (Aug. 14, 1975): 432–444, p. 435.

Digitized by Arya Samaj Foundation Chennai and eGangotri lodged: time travel seems to entail the possibility of the existence of a logically pernicious self-inhibitor. The objection is reminiscent of the argument against tachyons. Earman asks us to consider a rocket ship that at some space-time point x can fire a probe that will travel along a timelike loop into the past lobe of x's light cone. Suppose the rocket is programmed to fire the probe unless a safety switch is on and that the safety switch is turned on if and only if the "return" of the probe is detected by a sensing device with which the rocket is equipped (230-232). Is the probe fired or not? The answer is that it is fired if and only if it is not fired, which is logically absurd. Again. this contradiction does not suffice to show that time travel per se is impossible. Rather the whole situation is impossible, and this includes assumptions about the programming of the rocket, the safety switch, the sensing device, and so forth. But, although the contradiction could be avoided by giving up some of these assumptions, Earman suggests that we have good evidence that rockets can be so programmed. Earman concludes, "Thus, although we cannot exclude closed timelike lines on logical grounds, we do have empirical reasons for believing that they do not exist in our world" (232). His conclusion may be strengthened: it is not just the feasibility in our world of such rockets which generates the paradox; so long as such machines are nomologically possible, the contradiction could arise. Given the nomological possibility of such machines, then, timelike loops must be nomologically impossible if the contradiction is to be avoided. The conclusion would therefore appear to be similar to that in the tachyon case: that, although time travel is logically possible, there are no nomologically possible accessible worlds in which time travel can occur.

Paul Horwich has, however, disputed Earman's reasoning, claiming that he invalidly infers that, since the various assumptions are logically incompossible and since the rocket, safety switch, and so forth are physically possible, therefore timelike curves do not exist (440). But there could exist timelike curves in the actual world or in any physically possible world in which the rocket, switch, and so forth do not exist. Letting p = "The rocket, probe, safety switch, and so forth exist and function properly," q = "Timelike loops exist," and r = "The probe is fired," Horwich's argument appears to be that the following reasoning, which is Earman's, is invalid:

5. (i)
$$(p \cdot q) \supset (r \equiv \sim r)$$
 P
(ii) $\Diamond p$

²⁰ I am indebted to William Hasker for many interesting discussions of this issue.

(iii)
$$q$$
 p ADJ (iii), (iiii) (v) $(\lozenge p) \cdot q \to (r \equiv \sim r)$ (i) $(\lor p) \cdot q \to (r \equiv \sim r)$ (i) MPP (iv), (v) $(\lor p) \cdot q \to (r \equiv \sim r)$ (viii) $q \to \lozenge (r \equiv \sim r)$ CP (iii-vi) RAA

The problem is that (v) does not follow modally from (i). Although the conjunction of p and q implies an absurdity, the conjunction of q with $\Diamond p$ implies neither a contradiction nor even the possibility of a contradiction. In other words, timelike loops can exist in any world in which such rockets, switches, and so forth are possible but never in fact exist or function correctly; similarly for tachyons and the tachyonic antitelephone.

The opponent of time travel (and tachyons) has thus apparently committed precisely the same fallacy as the theological fatalist, and the response to them has the same form. The opponent of fatalism asserts that from God's foreknowledge of a future contingent proposition it follows, not that the future event cannot occur but only that it will not occur; the proponent of time travel maintains that from the fact that timelike loops exist it follows, not that such rockets cannot exist or function properly, but only that they do not exist or function properly. Further, the opponent of fatalism maintains that, if the contingent ent were not to occur, then different propositions would have been true and God's foreknowledge would have been otherwise; the proponent of time travel contends that, if such rockets were to be built and function properly, then the timelike loops would not exist. Thus, the two situations seem quite parallel.

IV. TACHYONS, TIME TRAVEL, AND THEOLOGICAL FATALISM

Now I must confess that, whereas the argument of the opponent of theological fatalism seems entirely plausible, the same argument in the hands of the proponent of time travel (and, implicitly, of tachyons) runs strongly counter to my intuitions. One might imagine a world, for example, in which all the technology and even the blue-prints for the rocket, probe, and so forth exist and in which timelike loops exist. It seems bizarre to claim that, while the rocket could be built, so long as no one in fact builds it, the loops can exist without the possibility of a contradiction's arising. Moreover, it seems very strange to claim that, were the rocket and so forth to be built, then the timelike loops would not exist. Suppose a team of rocket scientists took out the blueprints of the devices and decided, "Let's build them!" What is going to stop them? Horwich's response that to ask such a question is simply to ask why a contradiction does not come

true might fail to assuage one's suspicions that something is amiss here. Something must prevent the rocket's being built or a contradiction will arise; if the rocket and so forth are constructible, a contradiction would seem to be generable, which is absurd. Or again, we might imagine a world in which the rocket, probe, and so forth do exist and in which time travel occurs regularly. But each attempt to generate the self-inhibiting situation is frustrated by a series of accidents, which prevent the devices from functioning properly. But why do they always go wrong? Or worse, why do things not go wrong whenever the probe travels the same loop when no safety switch is used, but go awry whenever the switch is employed? Horwich confesses that he does not know the answer, but he believes that there is no reason to think an answer is impossible. This confidence might strike one as a somewhat unwarranted optimism. Finally, we might imagine a world in which time travel along timelike loops is a regular affair and in which the rocket, switch, and so forth not only exist, but would function properly if they were used. But in fact nobody uses them. Indeed, the commander of every time vessel may instruct his new recruits, "Do not activate the probe and the safety switch with the sensing device; otherwise the timelike loops along which we travel would not exist." Obeying his command, the new recruits like the rest of the crew are careful not to activate the devices, lest the loops should not exist. But does the very structure of space and time thus depend on the obedience of callow, young recruits to their commanding officer?

Nevertheless, it must be admitted that I have been somewhat unfair to the proponent of time travel in my illustrations. When we consider a world, we take into account not merely the history of that world up to some time t_n but rather its whole history. In any world containing timelike loops, the envisioned rockets never exist or function properly. It is not as though at t_{n+1} someone might build the devices and so cause the loops that had existed to fail to exist. Nor is it being claimed that the structure of space-time is dependent upon human decisions. Rather the point is that, since p and q are logically incompossible, their corresponding states of affairs never both obtain in any world. If one obtains, the other does not. If the other did obtain, then the one would not. To ask why is, as Horwich says, merely to ask why contradictions are not true. To think that in this case a contradiction is possible seems incorrectly to presuppose that time travel involves changing the past, an error analogous to the assumption, frequently made by theological fatalists, that one's freely choosing to do other than one does would involve changing God's foreknowledge. If the probe is seen to be returning though the

safety switch is on, the space travelers know that the switch is going to be turned off or malfunction in some way so as to permit the launching of the probe. If the switch is off, they know it or the probe is malfunctioning. Should they decide not to launch the probe after all, for some reason or other (malfunction, change of mind, disobedience to the commander) the probe will be sent anyway (and they no doubt realize this). Otherwise it would not be seen to be returning. For one cannot change the past.

I think that the sense of discomfort which the time travel case (like the tachyon case) evokes but which the case of divine foreknowledge does not elicit is due to the absence in the former case of a lack of a relation of conditionship (in Roger Wertheimer's sense²¹) between the existence of the time loops and the construction and functioning of the rocket. What is at issue here is a piece of counterfactual reasoning on the part of the proponent of time travel:

6.
$$p \longrightarrow \sim q$$

7.
$$p \cdot \sim q \longrightarrow r \equiv r$$

8.
$$p \longrightarrow r \equiv r$$

The reasoning is valid and purports to show that, if the rocket and so forth were to exist and function properly, then the probe would be fired iff it were fired, since no timelike loops would exist in such a world. The truth of (6) appears to depend at any point in time upon a special resolution of vagueness which permits backtracking counterfactuals, that is, counterfactuals in which the truth of the antecedent implies some adjustment of the past. In such a case the closest possible worlds to the actual world are not those in which the past is preserved inviolate, but in which some feature of the past is other than in the actual world in order that some overriding feature of the actual world might be preserved as much as possible. It is a highly disputed question as to when a special resolution of vagueness between worlds is warranted. It seems to me, however, that a special resolution is permissible when a relation of conditionship obtains between the state of affairs described in the antecedent of the counterfactual and that described in the consequent. Where this is lacking, the burden of proof would seem to lie on him who maintains that a special resolution is to be employed rather than the standard resolution of vagueness. Hence, for example, it seems true that

^{9.} If it were the case that Lincoln was assassinated and I can possibly eat ice cream, then were I to do so, it would be the case that Lincoln was assassinated and I eat ice cream.

²¹ See his "Conditions," this JOURNAL, LXV, 12 (June 12, 1968): 355–364.

Digitized by Arya Samaj Foundation Chennai and eGangotri
Here Lincoln's death and my eating ice cream are totally unrelated, and so whether or not I eat does not affect Lincoln's death. Analogously, the construction and proper functioning of the rocket have no effect upon the structure of space-time. Hence, if the timelike loops exist and the rocket and so forth are possible, then it seems that it would be true that, if the rocket were to exist, both the loops and the rocket would exist, which results in a self-inhibiting situation. But, since it is impossible that, were the rocket to exist and function properly, then both it and the time loops would exist, it follows that it must be impossible for the time loops to exist and the rocket to be possible. Since the rocket is possible, necessarily the time loops do not exist.

The crucial difference between these two cases, however, is that, although both lack a relation of conditionship between the earlier and later states of affairs, the time-travel case involves contradictory states of affairs, which the other does not. A backtracking counterfactual is therefore required in the time-travel case, not because the time loops are conditioned by later events, but because the envisaged situation does not obtain in *any* possible world; that is, there simply is no world in which both states obtain. The closest worlds to the actual world in which the rocket exists and functions properly must be worlds in which time loops do not exist. Therefore, a backtracking counterfactual is here in order, even under the standard resolution of vagueness and in the absence of any relation of conditionship between antecedent and consequent, despite the feeling of disquiet with which one is left.

This inquietude can, however, be considerably diminished by an analysis of one of the logical properties of "within one's power." Is the notion "within one's power" closed under entailment? That is to say, is

10. If (i) p entails q, and (ii) S has the power to make p true at t, then S has the power to make q true at t.

true? Joshua Hoffman and Gary Rosenkrantz have argued convincingly that it is not.²² For example, although it may be within my power to bring it about that

11. Some rocket ship is red.

is true, and 11 entails

12. Some rocket ship exists.

²² Joshua Hoffman and Gary Rosenkrantz, "On Divine Foreknowledge and Human Freedom," *Philosophical Studies*, XXXVII (1980): 289–296.

it may not be within my power to make (12) true. Therefore, power is not closed under entailment. Alfred Freddoso hopes to rectify the deficiency revealed by this important insight by requiring that p and q be logically equivalent. That is to say, he defends

10'. If (i) p is logically equivalent to q and (ii) S has the power to make p true at t, then S has the power to make q true at t.

Although he provides no justification for 10', he considers it "impeccable." ²³

But it seems to me that 10' may not be so flawless after all. For consider a situation such as that envisioned in Newcomb's paradox:²⁴ a being guesses in advance whether I shall choose one of two boxes B₁ or B₂. My choice has absolutely no influence on his prediction, nor is his forecast the result of precognition: it is pure guesswork. Let us, however, suppose that the predictor is infallible, or essentially inerrant. If follows that

13. I choose $B_1 \equiv$ The being predicts that I choose B_1 .

But, although it is within my power to choose B_1 , it is not within my power to bring about the being's prediction; for the problem conditions guarantee that the being's prediction is entirely outside my control. Therefore (10') is false. Now consider another scenario in which the notion of precognition is admitted. In this case the being cannot fail to predict my choices correctly because he has infallible precognition. So in this case, too, (13) is true. Here, however, it appears that it is within my power to bring about the being's prediction as well as my choice, since my choice determines his precognitions. But what about what lies within the being's power? It is within his power to predict that I choose B_1 , but it is not within his power to bring it about that I choose B_1 . So, once again, (10') is false. No doubt these cases are exotic, but then again power over the past is an exotic subject, and the cases have obvious relevance to the question at hand.

The above cases suggest that what is missing from (10') is some mention of the relation of conditionship between p and q. Only if p is a condition of q in Wertheimer's sense can one be guaranteed that, by having it within one's power to bring it about that p, one also has it within one's power to bring it about that q. Accordingly, I should replace 10' with

Robert Nozick, "Newcomb's Problem and Two Principles of Choice," in N. Rescher, ed., Essays in Honor of Carl G. Hempel (Boston: Reidel, 1969), p. 132.

²³ "Accidental Necessity and Power over the Past," *Pacific Philosophical Quarterly*, LXIII (1982): 64.

10*. If (i) p is logically equivalent to q, and (ii) S has the power to make p true at t, and (iii) q is a consequence of p, then S has the power to make q true at t.

Hence, even though it is true that

14. The rocket, probe, safety switch, etc., function properly

Time loops do not exist.

and, even if space cadet Jones has it in his power to bring it about that the first half of this equivalence is true, it does not follow that he has it within his power to determine the structure of space and time. All that follows is that Jones exercises his above power in worlds in which there are no time loops and that in worlds in which loops exist Jones never exercises his power. There is a sort of logical parallelism here without any relation of conditionship, and so rejection of the self-inhibitor argument does not imply embracing counterintuitive notions of power.

V. CONCLUSION

In conclusion, I think it is clear that the problems that confront the philosopher of science and the philosopher of religion respectively can turn out to be very similar and that interaction between the two can lead to some helpful insights for both. In the present case, the argument of the opponent of theological fatalism bears striking resemblance to the argument of the proponent of tachyons and time travel. They agree that past states of affairs may obtain which are logically incompatible with some envisioned action and yet insist that such an action is still possible because, if it were to be taken, then the past states of affairs would not have obtained. This is initially disquieting, since in the one context the argument seems quite plausible whereas in the other the results seem counterintuitive. This inquietude can, however, be alleviated, I have argued, by positing the presence of a relation of conditionship in the case of divine foreknowledge, which makes it reasonable to ascribe to a free agent the power to determine partially what God foreknows, a relation which is absent in the cases of tachyons and time travel, so that in these cases one has no power over the past.

Westmont College

WILLIAM LANE CRAIG

BOOK REVIEWS

Religious Experience. WAYNE PROUDFOOT. Berkeley, CA: University of California Press, 1987. xix, 263 p. Cloth \$32.00.

This book deals with Western discussions of religious experience from F. D. E. Schleiermacher onwards. The author treats the notion as a crucial one in reflections on religion in the past two centuries. After Kant, it was no longer easy to ground religious belief on metaphysics; and, in the turbulent aftermath of the French Revolution, the sanctity of institutional religion was even further undermined. Between science and ethics and beyond aesthetics, however, there was a niche for religion. Religious language was seen as essentially expressive—of religious affections, and above all of the feeling of absolute dependence. Proudfoot goes on to discuss various approaches to interpretation: he thinks of the human and social sciences as essentially concerned with interpretation; but this effort at verstehen does not preclude explanation. So later on he is critical of ways in which accounts of religious experience interpret it in a manner precluding explanation in some other terms (he is thinking especially of Schleiermacher, Rudolf Otto and D. Z. Phillips). In ch. III, he brings social-psychological theory to bear on emotions (including therefore religious ones), notably Stanley Schachter's twofactor theory of emotion which distinguishes physiological arousal from cognitive appraisal. In ch. IV, he discusses the question of a core of mystical experience, and follows mainly James' and Stace's characterizations. The authoritativeness of an experience, he argues, derives from the description under which the mystic identifies the experience. In ch. v, Proudfoot goes on to explicate the notion of religious experience, with "the explicit recognition that it is our phrase and has a rather parochial location in modern Western culture." His conclusion is that, to be religious, the experience must be inter alia, explicable in terms of the doctrine of the given tradition. So the very notion, according to Proudfoot, is supposed to incorporate an antireductionist aspect. In ch. vi, on explanation, Proudfoot pursues a similar theme. The protective strategies of Schleiermacher and Otto are ways of evading a call for naturalistic explanation. Such an explanation has to deal with what is identified under a description, and here we must incorporate the subject's belief that it derives from some transcendental object. Thus, if we ask for an explanation of a given person's experience (and he cites among others the case of

0022-362X/88/8503/0151\$00.50

© 1988 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri

one Steven Bradley), "we would need to know something about Methodist revivalism in early nineteenth-century New England, about the particular meeting he attended earlier in the evening, and about the events in his life up to that moment" (223).

This summary does not do justice to the detailed character of much of the discussion. The general drift of the book is, from the critical angle, justified. But the mode of approach does seem to me to be culturally limited, as is common among philosophers of religion. It is true that there are some mentions of Buddhism and some other Eastern religions. But there is little attempt to break out of the earlier Western tradition and to attack the question of the nature and types of religious experience on a more scientific basis. For instance, you cannot decide whether there are similarities of experience between traditions by citing such outdated figures as William James, who of course wrote wonderfully for his times, but did not have the benefit of the very great increase in our knowledge of the differing religious traditions. Proudfoot does not, either, much explore the via media, that is, the notion that we can exercise epoche or so-called bracketing, and so give a proper sense of the experience labeled "religious" without presupposing anything about the validity or otherwise of those experiences. He does not-to put matters differently-really explore the question of how we can frame methodologically agnostic hypotheses which could be tested. For there are more issues than of the kind: "What is the explanation of Stephen Bradley's conversion?" There are issues concerning the transcultural identity of types of religious experience, and whether religious experience has some kind of "autonomous" force or can be explained by other elements in human psychic development, etc. He does not even fully explicate what reductionism (a notoriously slippery term) means.

The fact that typically, or always, people delineate their experiences by reference to authoritative claims made within their own traditions need not mean that we cannot describe the essential characteristics of the experience crossculturally: though this would require some hermeneutical sophistication, seeing beyond, so to say, the surface of the words. Nor by the same token should we be precluded from investigating the varieties of religious experience (which is in my view a genus, and its species can be exhibited in a crosscultural framework). These more empirical investigations may mean that we come to use the expression 'religious experience' in a new way, going beyond the characterizations included in Schleiermacher, James, and Otto. It becomes a term of art for comparative purposes.

There is an issue about the relation of religious experience to

perception which needs exploring. Proudfoot emphasizes that a noetic quality is incorporated into the concept of religious experience, as it is into perception. But, in the former case, we do not, as investigators, wish to endorse "the claims that are constitutive of those experiences." The difference, then, between perceptions and religious experience is that the observer's identification of a perceptual experience does, and his identification of a religious experience does not, imply an endorsement of the claims assumed by the subject. I doubt this contrast. It is true that, when we use the verb "to see" and its cognates, we imply the truth of the claim embedded in the experience. It is convenient to do this. It is true also that, if we put in 'he seemed to', we implicitly cast doubt on the truth of the claim. But we know very well that a mistaken experience and a veridical one can be phenomenologically identical (for how else could we come to be deceived by our senses?). So it is always open to us to use some form of words to convey the fact that we are keeping open matters of truth claims about the objects or foci of experience, and concentrating on the qualities of the experiences themselves. So we can use bracketing of a sort equally in both cases. This reopens the question of whether we might substitute for Otto's account of the "holy" something which would not assume certain concepts and beliefs (of a broadly theistic nature). We might substitute, then, a cleaned up phenomenology which might convey such noetic and other qualities as the numinous experience possesses without building in a philosophical or Christian-theological claim. Proudfoot is quite right to detect the apologetic note in philosophers' appealing to religious experience: but the empirical study of religions, which is highly relevant to the philosophy of religion, can and has to dispense with that apologetic note. This possibility seems to be left out of Proudfoot's discussion, precisely because he approaches religious experience from the history of the subject through Schleiermacher, James, and Otto. But who do these dead men think they are? They had not yet fully grasped the possibilities of a more value-free type of exploration of humankind's religions. To return to a phrase I have used above, between the methodologically theistic approach and the methodologically atheistic approach there lies the methodologically agnostic one. Ultimately this seems to be what appeals to Proudfoot; but this book is largely aimed at bringing out the first view implicit in views of religious experience among philosophers of religion, and so is only a prolegomenon to a fuller treatment of the subject.

The book is written clearly and with economy of language. I think Proudfoot is right in thinking that his thesis has implications of a deep kind for the discussion of the relation between religion and science. Protective compartmentalization breaks down. There is perhaps an equally important relation between discussions of religious experience and the social sciences. The differing kinds of doctrine and religious practice are to some degree tied to varying emphases in the traditions or subtraditions on differing kinds of experience and feeling. So a phenomenology of religious experience has explanatory value (e.g., mystical experience in the narrow sense seems to conduce to some form or other of the *via negativa*) in the doctrinal dimension of a tradition. Religious factors in this and other ways have a certain human autonomy, which means that they become social forces. Here the question of their rationality or truth is irrelevant except in so far as some ideas can be weakened among certain classes, e.g., educated middle class people in some countries, if they do not match the criteria of truth commonly employed among those

But that takes us further afield. The book is well done and stimulating, if somewhat restricted in its purview.

classes. All this implies the importance of religious studies among the

NINIAN SMART

University of California/Santa Barbara

social sciences.

Bentham. ROSS HARRISON. Boston: Routledge & Kegan Paul, 1983. xxv, 286 p. Cloth \$34.50, paper \$14.95.

Ross Harrison's study makes Bentham interesting and accessible. Harrison guides the reader through a corpus of daunting scope and discouraging condition. His aim is primarily interpretative (vii), but his method is not (nor could it be) purely expository. Harrison develops his interpretations by working through philosophical puzzles that Bentham faced or left for posterity. This helps Harrison to identify Bentham's motivations, to reconstruct his doctrines, and to display a systematic coherence in Bentham's work.

The nine principal chapters in this book cluster in groups that deal with the themes of fiction versus reality (chs. 2–4), fact versus value (chs. 5–7), and public versus private ethics (chs. 8–10). I shall suggest how Harrison's method in the first of these three sections illuminates his subject, and then confine myself to an even briefer summary (with just a few critical comments) on the rest.

Among the legal fictions that first captured Bentham's attention were "wilful falsehoods" that courts accepted in order to expand their jurisdiction or modify the application of established legal doc-

0022-362X/88/8503/0154\$00.50 © 1988 The Journal of Philosophy, Inc.

trines. One might object to the use of fictions out of a commitment to truthfulness; but Harrison observes that, as Bentham was committed more firmly to utility than to veracity, we should expect his objections to be grounded on their disutility. But then one must consider William Blackstone's view of legal fictions as useful devices enabling courts to overcome legal obstacles to achieving justice.

Harrison links the issue with Bentham's attitude to common law generally. The common law system relies on doctrines that receive their first official acknowledgement when they are applied by courts in order to decide specific cases. Bentham saw common law as judge-made law, which establishes legal norms retrospectively. This not only usurps the authority of the legislature but, more importantly, undermines clarity and certainty for all those ordinary citizens who cannot assume that their legal liability will be limited to whatever was provided by law when they acted. Under a common law system, we can be subjected to legal sanctions after the fact. Legal fictions are similar to common law doctrines not simply because they are judicial creations which substantively affect the law, but also because they make it more difficult for ordinary people to understand the law. Bentham held that they reinforce the great, oppressive power of lawyers and judges. He seems to have regarded it as obvious that the costs of such legal arrangements far exceeded the benefits.

The second category of fictions, and the one of primary concern to Bentham, contains such things as rights and obligations. These are fictions because, he believed, there is no *thing* to which, for example, the substantive 'obligation' might refer. Bentham nevertheless believed that these locutions make a kind of sense. He distinguished "fictional entities" such as obligations, which he took to be unavoidable implications of our common language, from "fabulous entities" such as golden mountains, which are simply creatures of our imagination.

Bentham developed a theory of linguistic fictions as part of his lifelong project to illuminate and reform the law. Along with a deontic logic, one of his most impressive philosophical innovations is a method of analysis called "paraphrasis," in which a substitute sentence is to be provided for one that incorporates a term that, taken literally, refers to a fictional entity. One would clarify a term like 'obligation' by translating the ascription of an obligation into a sentence about "sensibles," which are *real* entities in Bentham's universe.

Harrison's interpretation of this theory is directed by his exploration of such problems as explaining how a sentence that succeeds in clarifying could be the proper translation of one that requires clarifi-

cation and, given their differences, how an appropriate translation could be identified. Another serious problem that receives extensive treatment by Harrison derives from the fact that Bentham accepts linguistic fictions as unavoidable and even useful. He tolerates talk about legal rights; how then can he distinguish them from, so as to justify his condemnation of talk about, "natural" rights?

Just as legal rights are seen by Bentham as creatures of legal obligations, and thus of ordinary laws, so "natural" rights can be seen as presupposing "natural" laws. Harrison observes that Bentham cannot simply reject the very idea of a "natural" law as nonsense, because his own principle of utility amounts to such an idea. Bentham can argue, however, that one's having or lacking a legal right makes a difference that having or lacking a "natural" right does not. To have a legal right is to be under the protection of ordinary laws, which have the character of orders backed by threats from the sovereign political power within the community. Talk about "natural" rights has no such implications; it implies no real protection. The ascription of a "natural" right is really a claim that there ought to be a law creating a legal right protecting the corresponding interest. Bentham held that talk about "natural" rights misleads people about the protections they enjoy under ordinary law. This could explain his condemnation of such talk.

Harrison argues that Bentham's attack on "natural" rights was motivated even more fundamentally by the importance within his system of the fact-value distinction. Ordinary laws are matters of fact, but "natural" laws are not; such talk is "normative," not "descriptive"; and proponents of "natural" law "confuse questions of fact with questions of value" (103). Ascriptions of legal rights can be analyzed by paraphrasis, because they are matters of fact; but ascriptions of "natural" rights cannot be so analyzed, because they are really normative judgments, to be understood as claiming that certain rights ought to exist. For this reason, too, talk about "natural rights" involves bad rather than useful fictions: it pretends to be, but is not really, factual. In this reading (and elsewhere, e.g., in denying that Bentham embraced a reductionistic naturalism), however, Harrison ignores the possibility that normative language does not rule out factual content and that Bentham might best be understood as a nonreductive moral realist.

Harrison emphasizes that Bentham's system is built upon two principles: not just the principle of utility, "that pleasure ought to be maximised and pain minimised," but also the psychological principle "that people act with the motive of maximising their own pleasure and minimising their own pain" (109). This accounts for Bentham's preoccupation with designing institutions "so that each person following his own self-interest will in fact do what will result in general utility" (115). Bentham realized that the point applied to officials. Its crucial importance led him to concentrate on the problem of motivating officials so that, as they inevitably pursue their own interests, they will also serve the general welfare.

Harrison appreciates that what passes for psychological hedonism can reduce to a claim that is either trivial or false. According to Harrison, Bentham held that, as a matter of fact, people sometimes act unselfishly, but so infrequently that prudent institutional design must assume the contrary. Bentham nevertheless believed that pleasures and pains are the fundamental motivators of human action, but he seemed uncertain that the hedonic calculations required by his principles were in fact possible.

Harrison says that Bentham's "indirect proof" of his principle of utility amounts to "a fairly powerful argument" (185). Bentham argues that his principle alone renders moral questions decidable by public, empirical tests. Harrison claims that Bentham thus shows that his principle uniquely satisfies a basic requirement for fundamental moral principles. Unfortunately, Harrison fails to defend the claim that there is such a requirement or, more importantly, to appreciate that innumerable competing principles (e.g., one requiring equal welfare) satisfy it (or fail to do so) for precisely the same reasons.

Bentham gradually came to believe that popular control in the form of representative democracy was the best insurance that institutions would serve the general happiness. But even this solution could not implement itself. To achieve popular control, one must persuade rulers that democracy is the only alternative to revolution.

Harrison suggests that utilitarian institutions would respect traditional moral distinctions and meet many standard objections—or at least that this is how Bentham imagined that his theory would work in practice.² Harrison argues further that Bentham's use of the criterion of utility commits him to act rather than rule utilitarianism. But one might have reservations for two reasons. First, Bentham applied the criterion to whatever could be said to have utility, and his utilitarianism might well be more complex than the standard categories

¹ But, although he saw through the objections, he failed publicly to support suffrage for a good part of "the greater number," namely women.

² Harrison believes that Bentham solved the free-rider problem, apparently because he does not see that utility and not merely self-interest can argue for free-riding.

Digitized by Arya Samaj Foundation Chennai and eGangotri

envisage. Second, as Harrison emphasizes, Bentham was more concerned with the shape of institutions than with moral guidance for the conscientious individual. His book ends in fact with an interesting discussion of the difficulty Bentham has in fitting a utilitarian "deontology" into his psychological framework.

This is a good, useful, and welcome book. Harrison is a sensitive interpreter with a helpful grasp of an enormous body of material. His occasional suggestion that he has successfully defended some of Bentham's substantive views should in fairness be discounted. It is, as Harrison says (viii), "a book about Bentham, not a work of modern philosophy."

DAVID LYONS

Cornell University

Foundations of Space-time Theories: Relativistic Physics and Philosophy of Science. MICHAEL FRIEDMAN. Princeton, NJ: University Press, 1983. xvi, 385 p. Cloth \$35.00.

Michael Friedman's imaginative, thoughtful, and carefully worked out study of the fundamental philosophical issues at the foundations of contemporary space-time theories in physics constitutes a very major contribution to the literature in this field. Since Friedman correctly understands that the resolution of the most important of these philosophical questions can only be done in a broader context of a foundational philosophical investigation into the nature of theories in general, the book should be given the careful attention of not only those interested in the philosophy of physics or of space and time, but of those concerned with more general issues of realism and conventionalism in our naturalistic account of the world.

Much of the book is devoted to expounding the fundamental structure of the most important space-time theories of physics in the four-dimensional and modern geometric formulation which presents the alternatives in the most perspicuous way. The exposition is self-contained, with the fundamental mathematical devices being introduced as needed, and provides a marvelous introduction to this material for the reader who has the modest mathematical sophistication needed to grasp the formalities. Prerelativistic theories such as Newton's, the neo-Newtonian theory that eschews absolute rest, Newtonian gravity in both Newton's original force-field version and

0022-362X/88/8503/0158\$00.50

© 1988 The Journal of Philosophy, Inc.

in the modern geometrized version of Andrzej Trautman, and the relativistic theories of the special theory and the general theory all find their place in this framework and, in it, the inter-relations between them can be clearly laid out.

Within the framework, important questions such as the nature of a so-called "relativity principle," versions of "equivalence principles" of varying generality and strength, and the inter-relation of a theoretical structure with the idealized observational data that ground it are all given extensive and perspicuous treatment.

The major philosophical aim of the work is to defend a "realist" view of space-time ontology against its critics. Relationists deny the reality of space-time itself taken as a constituent of the world over and above the ordinary material objects and their features. Conventionalists deny the uniqueness of any one theory of space-time being the "true" account of the world, arguing that what we take the geometry of space-time to be is a matter of our choice and convenience. Since the realist denial of relationism posits a world of "unobservables" and their features, and since it insists that one only of a number of alternative accounts equally compatible with "all possible observations" is the true account, the task of the realist includes the job of assuaging the inevitable skeptical doubts that, even if such a unique, ontologically vigorous, space-time theory is correct, we could ever know which such theory is the correct one.

Against the relationist who would give the apparently referential theoretical apparatus of space-time theories an, at best, "as if" or fictionalist reading, Friedman argues for an inference to theoretical entities and their features as providing "unifying explanations" of the observable phenomena. Insofar as the positing of the theoretical ontology serves this unifying explanatory role, it ought to be taken as genuinely referential.

The fundamental argument here is a curious and ingenious one owing its inspiration to William Whewell (on the consiliance of inductions) and to Richard Boyd and Hilary Putnam (on unifying theories by conjunction). The interpretation of a theory with its theoretical term taken as genuinely referential is contrasted with its Ramsey sentence surrogate, where the existential quantification in the Ramsey sentence can be contrued as ranging over abstract objects (such as numbers) and as asserting only the possibility of a "representation" of the observable data in an abstract structure, much as we can represent data relative to a coordinate system imposed on space-time which, in itself, reveals no "real" structure of the world. It is argued that two theories with the same theoretical

term can be conjoined into a new broader theory which has consequences that go beyond the mere conjunction of the consequences of the individual, now component, theories. But the conjunction of the Ramsifications of the theories has no such broader consequences, since each variable that has replaced a theoretical term is "trapped" in its own existential quantifier. Then it is argued that this fact about the historical evolution of theories shows us that a theory realistically construed can obtain greater confirmation (by cross confirmation of one component of the new broader theory by the data that originally supported only what is now the other component and by the joint consequences) than can the theories construed in the antirealistic, representational manner.

I find this particular version of an argument for realism from unification rather puzzling. If the two theories at one time shared a theoretical term, intended uniquivocally, then the proper Ramsey interpretation of the theoretical picture at that time was a single Ramsified theory with a single quantified variable filling in the gaps in the conjoined matrix of the Ramsey sentences of the two theories taken separately, and this would function in a confirmational vein exactly as does the conjoined theory in the realist view. It would hardly take "genius" to formulate the unified Ramsey sentence prior to the new evidential consequences of the unified theory coming in. The work of genius would have already been done in using the same term, intended with the same meaning, in the two prejoined theories.

It is also unclear to me how this model of realism will be able to generate all of the consequences Friedman wants from it, in particular how it will justify positing a fully substantivalist space-time account, complete with unoccuppied space-time locations. It is not at all clear to me that some alternative "thinner" account, say one positing basic space-time properties but eschewing locations, could not be formulated which would serve the same "unifying" role as would the "substantival" view. Friedman correctly points out that the usual Leibnizian "indiscernability of possible worlds" arguments against space-time substantivalism rest on posited symmetries of the models, symmetries which do not generally exist in general relativistic space-times. But more general Leibnizian arguments, in which space-time points, taken as entities in their own right, are "pushed" through the space-time structure, once again suggests that an excess of models has been proposed. Although a "structuralist realism" would still be a space-time realism, it would seem that such an account might satisfy all the constraints of being a unifying explainer which Friedman demands as well as would the richer substantivalist

account with its Leibnizian difficulties. It is unclear, of course, what such a "structural realism" would amount to, and unclear that it could do the job intended for it. But I think that the argument against this position and in favor of the usual substantivalist account —with its individuated space-time locations, occupied or not, as fundamental ontology—is independent of the general argument for realism Friedman offers here and must go beyond this general realist claim.

An argument for realism (and in the case of space-time theories for antirelationism) from some notion of the need for space-time and its structure itself to explain the phenomena might be constructed, but I remain skeptical, even in the light of Friedman's ingenious and carefully worked out account, that we will be able to show that a realistic theory is in any way better confirmed than its representational (or other eliminativist) alternative, even when we take the evolution of theories and the order in time of data and hypotheses into our confirmational account.

If relationism, according to Friedman, posits too few entities in the world, his argument against conventionalism is that the usual alleged "conventionalist" alternatives to our standard theories are too ontologically otiose, or, at least, too theoretically encumbered, to be methodologically acceptable.

Theories with Newtonian rest frames (as opposed to neo-Newtonianism which eschews them), aether theories (as opposed to special relativity), and flat space-time plus gravitational field theories (as opposed to general relativity), all suffer from having so much theoretical apparatus and so much "play" at the theoretical level that various fundamental parameters of the theory remain underdetermined by all possible data. We cannot tell which inertial frame is Newtonian rest or aether and we cannot observationally tell the difference between one world and another similar to it but suffused with a uniform gravitational force. And, surely, Friedman is right in thinking that our preference for the standard theory is its lack of arbitrary, empirically undeterminable and otiose elements when compared to its alleged "conventional" alternatives.

His model for this preference is one which combines a restrictive notion of theoretical equivalence with the positing of (a prioristic?) methodological rules of intrinsic preferability among theories. For Friedman, observational equivalence is not enough for full equivalence, lest we be led down the slippery slope to phenomenalism. In fact, he takes even very rigid conditions, such as interdefinability at the theoretical level, as insufficient for full equivalence, since he

Digitized by Arya Samaj Foundation Chennai and eGangotri

takes the Newtonian rest frame theories to be not only not inequivalent to neo-Newtonianism with which they are not simply interdefinable, but inequivalent to each other. Since no general account of meaning or meaning accrual is offered, I am not sure just what it would take for theories to be equivalent in Friedman's own view. Surely not exact verbal identity, for even he would, I hope, take electromagnetism with the words 'positive' and 'negative' interchanged as telling us the same thing about the world.

Given a set of observationally equivalent but distinct theories, we ought to accept that theory which is "methodologically preferred" to all the others. Given the "methodological defectiveness" of the usual "conventional" alternatives to the orthodox theories, we are to reject them in favor of the standardly accepted usual theory. What is the methodological defectiveness of the alternatives, now viewed not as conventional alternatives at all but discredited genuine alternatives? That there are too many of them all on a par with one another both observationally and structurally and that they are distinguished from one another only by a parameter whose value is empirically undeterminable given their own posits about the world. Fair enough. But what rationalizes our belief that the usually preferred alternative is more likely to be the *true* account?

Friedman does not offer us the rationale behind adopting this methodological rule we would expect from a nonskeptical realist. He does tell us that his rule is a rule which leads us to dispense with the so-called "conventional" alternatives in favor of the usual account and yet does not lead us further down the slope to positivism or skepticism about unobservables in general, and that it conforms to scientific practice. The general principle behind it is that of parsimony. We are to reject all the theoretical entities we can until we get to those requisite for explanatory unification. If we could accept the arguments for realism offered earlier, and accept that they necessitate in our explanatory scheme the ontology of the usual theories, then accepting such a general rule would at least focus our remaining skeptical doubts on Ockham's Razor like principles in general. If we are skeptical of the earlier arguments for realism, then this parsimonious stance is likely, without the "bound from below," to lead us once again toward positivistic reactions.

To skepticism directed at the methodological rule, Friedman offers mostly the familiar claim directed at the inductivist positivist that, without some methodological principles, we could no more generalize from observed to unobserved than we could from observ-

able to unobservable. Responding to Hans Reichenbach's distinction between genuinely alternative theories distinguished by inductive simplicity and the alternatives in these cases distinguished by "mere descriptive simplicity," he responds by challenging the distinction. To be sure, the fact that an inductively arrived at generalization can be refuted by possible experience (the 'can' being subject to familiar variations on the interpretation of the modality) does not in itself offer a rationale for the rule of induction we choose. But, if we once accept the view, as Friedman does, that the unobservables of our theory are genuine "in principle" unobservables, then does not the claim of Poincare repeated in Moritz Schlick and Reichenbach—that the choice made between theories differing only in the realm of the unobservable so lacks the impact of a "real" choice between hypotheses, which differ in the observable outcomes they predict, as to be denied being called a choice of anything but mode of expressionstill seem persuasive?

In any case, Friedman's principles for thinning ontology are of a more familiar vein than those invoked to put an end to the eliminativist program at the floor of entities genuinely needed for explanation. And to me they are more intuitively persuasive.

What we need, of course, to be satisfied with the realist program that Friedman offers, would be to go beyond his subtle mode of delineating what kinds of principles we would need to justify the invocation of all and only the theoretical entities he wants to invoke (and which we commonly invoke in our accepted space-time theories), and to back up these principles in a persuasive way. In the direction of guaranteeing enough theoretical entities, we need either to fill out the "consiliance of inductions" argument until it is persuasive or to infer to the unobservable in some other way, say by figuring out what explanation might be over and above lawlike subsumption and showing how space-time is, in that way, genuinely explanatory. In the direction of slimming our ontology down without sliding into an ontology without unobservables at all, we would need some general account of meaning in theory which would show us why we ought to accept, with Friedman, the view that empirical equivalence and even interdefinability are not sufficient for full theoretical equivalence. And we would need to find in our accent of theoretical meaning what would be sufficient for full equivalence of verbally distinct theories. Having done that, we would need to describe the general methodological principles of acceptance which, given our rigid notion of equivalence, allow us to dismiss all but one of our Digitized by Arya Samaj Foundation Chennai and eGangotri

empirically equivalent but contradictory theories. And, finally, we would need some way of rationalizing our belief that theories so selected are more likely than the others to be true in the correspondence sense of 'true' which adheres to a realistic notion of truth irreducible to warranted assertion.

Overall, Foundations of Space-time Theories is an extremely rich and important book. In addition to the probing critiques of relationism and conventionalism in space-time theories on which I have focused, the book treats a wide variety of the puzzles of space-time theories, disentangling many confusions and illuminating all the issues it touches upon. Relativity principles and symmetry principles, invariance and its contrast with covariance, and the range of posits that can be called equivalence principles, varieties of notions in which a theory can be "absolute," and, most importantly, the way in which an overall structure can be found which encompasses both the range of space-time theories and the systematic moves that constitute the essence of the modern revolutions in our understanding of space-time, thereby opening up a wide range of specific issues to a general philosophical treatment—are all treated with precision and clarity.

The book will constitute, for many years, essential reading for anyone who wants to pursue the philosophical issues at the foundations of space-time theories or to probe the issues of realism and conventionalism in our understanding of theories in general.

LAWRENCE SKLAR

University of Michigan

The Philosophy of W. V. O. Quine, vol. XVIII of The Library of Living Philosophers. Lewis Edwin Hahn and Paul Arthur Schilpp, eds. La Salle, IL: Open Court, 1986. xvi, 704 p. Cloth \$49.95, paper \$24.95.

The latest volume of *The Library of Living Philosophers* is a tribute to the work of W. V. O. Quine. The book consists of a forty-four page autobiography by Quine, and twenty-four essays on his work, each with a brief response; also included is a bibliography compiled by Quine himself, and a comprehensive index. The book is attractively

0022-362X/88/8503/0164\$00.50

© 1988 The Journal of Philosophy, Inc.

produced, and I noticed few misprints. It is sad to have to report, however, that the paperback edition which I read did not survive even one reading without beginning to lose a page here and there. A book of this size, which is also one that most purchasers will return to again and again, needs to be sewn, not glued.

The autobiography goes more deeply into logic and philosophy than does the longer The Time of My Life: An Autobiography, 1 and is quite striking. In particular, one sees how far logic and philosophy have come in the fifty-five years since Quine's professional career began. Quine taught himself logic by reading John Venn, Louis Couturat, Giuseppe Peano, and, especially, A. N. Whitehead and Bertrand Russell's Principia Mathematica. For the last forty years, at least, no student has had to have that much tenacity. The change is in no small part due to Quine's gift for expounding logic-from truthfunctions to set theory—in a rigorous yet lucid manner. Two of the essays discuss Quine's work in logic: an elegant essay by Joseph S. Ullian and a somewhat more critical piece by Hao Wang. To employ a distinction of Ullian's, Quine's contribution has been chiefly to the philosophy of mathematical logic, rather then to mathematical logic itself; his emphasis on clarity and rigor of exposition is surely relevant here, and thus of more than pedagogical significance. Philosophically, too, the landscape has changed dramatically, and Quine's contribution has been even greater. Here, of course, the notion of progress is more controversial than in logic. Few, however, will disagree with Wang's assessment that Quine has been, since the 1950s, the dominant figure in that strain of philosophy which looks for inspiration to Russell and to Rudolf Carnap.

In a volume of this sort, the choice of contributors is crucial. The editors, not surprisingly, have for the most part played it safe, and the majority of the contributors are eminent and well-established figures whose work is more or less within the analytic tradition. The individual essays which make up most of this volume are, as might be expected, a diverse lot. In subject they include the philosophy of mathematics, the philosophy of space-time, epistemology, holism, nominalism, and various aspects of the philosophy of language (as well as the essays on logic which have already been mentioned). The fact that all of these discussions are clearly relevant to Quine's work reminds us of his range. He has in fact put forward a complete worldview, a metaphysics, if you like. Another source of the diversity

¹ Cambridge, MA: MIT Press, 1986.

Digitized by Arya Samaj Foundation Chennai and eGangotri

of the essays is more regretable: they vary greatly in date of composition. Each has the date when it left its author's hands, and there is a gap of twelve years from first to last. Eight of the contributors had to wait over a decade before their work appeared (in this volume at least), and several more almost as long. One cannot help but think that they have a right to feel ill-used.

The general level of understanding of Quine's work demonstrated in the essays seems significantly higher than that of the essays in Words and Objections.² There are still places at which it seems clear, both to Quine and to the present reviewer, that an author has simply misunderstood Quine; but these places are comparatively few, and are usually at points where the issues are subtle or the texts difficult. To this extent, at least, there is progress in philosophy. I found three or four of the twenty-four to be not merely worth reading but worth pondering; I expect to return to them. Most of the rest seemed to me worth reading, at least for one with an interest in the given subject matter. Relative to the eminence of most of the contributors, however, the overall quality of the essays seemed to me disappointing. (These judgments are ones which I would expect to command general agreement; I do not say, however, that everyone would put exactly the same essays in these categories.)

Rather than attempt to consider one or more of the essays, I shall instead discuss Quine's replies. In general, they show all of the incisiveness and force which we have come to expect from this author. As is also typical, they are in some instances too brief: more would have been useful to ensure that the point got across. There is, as one would expect, a good deal of allusion to the author's other works—especially to the essays in *Theories and Things*, which was published after most of the contributions were written (in some cases Quine shows a rather optimistic faith that the contributor would not have held a given interpretation, or advanced a given argument, had that book been available to him). Nevertheless, the replies contain much that will be of interest, even to the most assiduous student of Quine's earlier works. A few points can be briefly discussed.

There is some enlightenment on Quine's doctrine of ontological relativity. According to this doctrine, the reference of a given term in a language is relative to the choice of a particular scheme of translation from that language into another. But the same is true of the language into which the first is translated; thus, some have thought that we were threatened with a regress, in which firm ground would

Donald Davidson and Jaako Hintikka, eds. (Boston: Reidel, 1969).
 Cambridge, MA: Harvard, 1981.

never be reached. Quine's response has been to say that the regress is ended by our "acquiescing in our mother tongue and taking its words at face value." 4 Clarification of this rather cryptic utterance emerges in the first essay in Theories and Things, and more in this volume. Remarks in Quine's "Reply to Roth," in particular, seem to me helpful in this regard, making it clear that to take words at their face value is not to understand them in some special way but simply to use them, without for the moment raising the question of their meaning or their reference. If that question is raised, it is raised relative to another language which is, in its turn, being taken for granted—used and not questioned. (Remarks by Robert Nozick, at the end of his essay, suggest a similar view.) Along with this clarification of ontological relativity goes an increased emphasis on this doctrine, and a consequent denigration of the importance of reference and ontology. More than before, it is clear that, for Quine, it is sentences and their truth which are primary; reference is a by-product (see the reply to Dagfinn Føllesdal, as well as to Paul A. Roth).

A second point concerns the much controverted notion of analyticity. Quine states "that the philosophically important question about analyticity and the linguistic doctrine of logical truth is *not* how to explicate them; it is the question rather of their relevance to epistemology" (207); their relevance, in his view, is of course none. This view was implicit in some passages in *The Roots of Reference*, but it is good to have it made quite explicit. The point is, moreover, linked with an important retrospective on "Two Dogmas of Empiricism." According to Quine's present view, it is the second dogma—reductionism—that is important; if this is abandoned, then analyticity (the first dogma) loses its general significance (see the reply to Geoffry Hellman).

Again, I was struck by Quine's emphasis (as against some commentators) on the importance of the distinction between the observational and the theoretical. He insists that observation sentences, unlike others, "do have their empirical content individually" (427; compare also the reply to Arnold B. Levinson). This is an explicit qualification to his holism. He argues that observation sentences have no analogues among judgments of value, and that this makes for a crucial difference in status between moral judgments and cognitive ones (see the reply to Morton White). This point is of general importance for an understanding of Quine's differences from the older American Pragmatists. Again, Quine sees' the major disconti-

⁴ Ontological Relativity and Other Essays (New York: Columbia, 1969), p. 49. ⁵ La Salle, II: Open Court, 1974.

Digitized by Arya Samaj Foundation Chennai and eGangotri nuity within a theory as occurring not, as some have thought, between concrete objects and abstract objects, but rather between observation sentences and theoretical sentences (see especially p. 402).

More briefly: it was news to me, at any rate, that Quine takes his rejection of necessity, including physical necessity, to imply the rejection of the distinction between law and accidental generalization (398). Finally, a constant theme of Quine's replies is the insistence on naturalism or realism as the crucial element in his thought (see especially Roger F. Gibson's essay and Quine's reply, and the replies to Harold N. Lee, Nozick and Hilary Putnam). Here, as elsewhere, Quine's replies do more than repeat the familiar; they give his readers new points to ponder.

PETER HYLTON

University of California/Santa Barbara

NOTES AND NEWS

The Tenth Annual Conference of the Cognitive Science Society will be held at McGill University, August 17-19, 1988. The conference will include major addresses and invited symposia on problem solving and reasoning, formal analysis of language, and the interaction between neurosciences and cognitive science. Submissions of papers, symposia, and posters are encouraged in any of the areas of cognitive science. Information may be obtained by writing the Cognitive Science Secretariat, 3450 University Street, Montreal, Quebec H3A 2A7, Canada.

The Sociedad Filosofica Iberoamericana (SOFIA) announces its first international conference, to be held in Tepoztlan, Morelos, Mexico, on August 8-12, 1988, on the topic "Information-theoretic Semantics and Epistemology." The participants will be Donald Davidson (Berkeley), Marcelo Dascal (Tel Aviv), Fred Dretske (Duke), Hartry Field (USC), Jerry Fodor (CUNY), James Higginbotham (MIT), Jaegwon Kim (Brown), Ernest Lepore (Rutgers), Brian Loar (USC), Barry Loewer (South Carolina), Christopher Peacocke (New College), Stephen Schiffer (Arizona), John Searle (Berkeley), Ernest Sosa (Brown), John Perry (Stanford), and Dennis Stampe (Wisconsin). All inquiries may be addressed to Enrique Villanueva, Apartado postal 22-423, Mexico 14000, D.F.

The Institute for the Advancement of Philosophy for Children at Montclair State College announces that the students in the 1988-89 master's program in Teaching Philosophy to Children will be eligible for assistantships made possible by a grant. The assistantships will provide recipients with a stipend and tuition waiver. The master's program allows students with a Bachelor's degree and a strong background in philosophy to obtain a Master's degree plus certification in less than one year. Applicants interested in matriculating in the program may obtain information by writing Ann Margaret Sharp, Institute for the Advancement of Philosophy for Children, Montclair State College, Upper Montclair, NJ 07043.

The Department of Philosophy of Rutgers University is pleased to announce the establishment of an annual international competition for the best essay written by an undergraduate majoring in philosophy. Initially funded for a three-year period, the deadline for the competition will be November 1st of each year. Submissions will be judged by members of the faculty of the philosophy department at Rutgers. Finalists will be announced on February 1st of the following year. The student who is judged to have written the best essay will be awarded a prize in the form of a credit account for books of his or her choice. Submissions should be sent to the attention of: Regina Mayer, Department of Philosophy, Rutgers University, New Brunswick, NJ 08903.

The Society for the Philosophy of Sex and Love wishes to receive papers for presentation at its meetings to be held with the Pacific Division of the American Philosophical Association (March 1989) and the Eastern Division of the APA (December 1989). Papers for the former meeting are due by September 15, 1988; for the latter, March 15, 1989. Papers may be on any topic related to the philosophical investigation of sex and love. Two copies may be sent prepared for blind reviewing to: Alan Soble, Phil. Dept., University of New Orleans, LA 70148.

The Department of Philosophy and the School of Law of the University of Miami will sponsor a conference on the topic "Rights and the Philosophy of Law," March 25–27, 1988, in Coral Gables. Speakers will include Carlos Alchourron (Buenos Aires), Eugenio Bulygin (Buenos Aires), Alan Goldman (Miami), Patrick Gudridge (Miami), Lars Lindahl (Lund), Neil MacCormick (Edinburgh), David Makison (Paris), Thomas Nagel (New York), Carlos Nino (Buenos Aires), Joseph Raz (Oxford), Eduardo Rabossi (Buenos Aires), and Roberto Vernengo (Buenos Aires). Information may be obtained by writing Risto Hilpinen, Phil. Dept., University of Miami, Coral Gables, FL 33124.

The University of Minnesota will host the Seventh Annual Meeting of the Group for Research into the Institutionalization and Professionalization of Literary Studies, April 20–23, 1989. Speakers will include Hayden White (Santa Cruz) and Evelyn Fox Keller (Berkeley). Papers for the various sessions and panels must be submitted by October 15, 1988. More information may be obtained by writing to David Shumway, Department of English, Carnegie-Mellon University, Pittsburgh, PA 15213.

The Seventh Annual Conference on Philosophy of Science, sponsored by the Center for Philosophy of Science of the University of Pittsburgh and sup-

ported by a grant from the R. K. Mellon Foundation, will be held December 9–10, 1988. The theme of the conference will be the ramification of evolutionary epistemology. In particular, papers will be welcome which consider the question of a validating role for evolutionary considerations in epistemic matters, either by way of particular cases or of considerations of general principle. Those wishing to participate in the program should send their submissions no later than May 15, 1988. Further information may be obtained from Nicholas Rescher, Center for Philosophy of Science, 817 Cathedral of Learning, University of Pittsburgh, PA 15260.

The Department of Philosophy of the University of Connecticut/Storrs is pleased to announce that Professor Eddy Zemach of Hebrew University, Jerusalem, will be a Visiting Professor in 1988–89. The Department is also pleased to announce the appointments of Leonard D. Katz as Assistant Professor and Diana T. Meyers as Associate Professor.

The Department of Philosophy of the University of Miami is pleased to announce that Risto Hilpinen has been appointed Research Professor of Philosophy, effective January 1988.

PRAXIS INTERNATIONAL

Edited by Seyla Benhabib and Svetozar Stojanović

In its first five years of publication, **Praxis International** has emerged as the voice of humanist, democratic socialism, providing a forum for the critical analysis of existing social systems, of their institutions and ideologies, of new social movements and of social theories. Distinguished contributors to past volumes have included J. Habermas, C. Taylor, A. Giddens, R. J. Bernstein, C. Offe, M. Markovic, A. Heller, R. Rorty, M. Jay, R. Falk, and M. Lowy. Forthcoming issues will discuss Foucault and critical theory; the USA; and neo-Aristotelianism.

April, July, October and January Subscription Rates Volume 7, 1987 Individuals: £16.50 (UK) £20.65 (overseas) US\$36.00 (N. America & Japan) Institutions: £47.50 (UK) £59.35 (overseas) US\$95.00 (N. America & Japan) Please send orders with payment to:

Iris Taylor Journals Department Basil Blackwell 108 Cowley Road Oxford OX4 1JF England

or to: Journals Department Basil Blackwell PO Box 1320 Murray Hill Station NY 10156 USA

Basil Blackwell · Oxford and New York

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 4, APRIL 1988

HUMAN NATURE AND SOCIAL CHANGE IN THE MARXIST CONCEPTION OF HISTORY*

CCORDING to Karl Marx's Theory of History, the fundamental explanation of the course of social change lies in facts that are in an important sense asocial, and, in one sense of the word, material.2 The relevant fundamental facts are asocial in that no information about social structure enters into their formulation, under the following understanding of social structure: a statement formulates a fact about social structure if and only if it entails an ascription to (specified or unspecified) persons of rights or effective powers vis-à-vis other persons.3 Examples of material facts in the present partly technical sense are the general one that human beings are able to sacrifice present gratification for the sake of greater future gratification, and the particular one that the productive resources available to European humanity in 1250 ensured that most labor in Europe would be agricultural labor.

KMTH offers an argument⁴ in which exclusively asocial premisses are used to support a grand conclusion of an asocial kind about the whole course of human history, to wit, that there has been, across that course, a tendency for the productive power of humanity to grow. This claim was called, in KMTH, the Development Thesis, and it was formulated as follows:

(1) The productive forces tend to develop throughout history (134).

As just indicated, the tendency posited by (1) is supposed not to be due to the character of social structures. If it obtained only because

² See KMTH, ch. IV, for an attempt to specify that sense.

^{*} We are indebted to Joshua Cohen, who commented extensively and acutely on a draft on this paper, and to John McMurtry, who suggested an improving clarifica-

¹G. A. Cohen (Princeton: University Press, 1978). This work will henceforth be designated KMTH.

³ This characterization of social structure is adapted from KMTH, p. 94. ⁴ KMTH, pp. 152/3. The argument is summarized in section 1 below.

of favorable social relations of production which were not themselves materially explained, it would then lack the autonomy with respect to social structures which KMTH assigns to it. One may, accordingly, enter the following alternative statement of the development thesis:

- (2) There is an autonomous tendency for the productive forces to develop.
- (2) will be called the Full Development Thesis.

Since productive forces are not unmoved movers, the autonomy here assigned to their tendency to develop is not an absolute one. The tendency's autonomy is just its independence of social structure, its rootedness in fundamental material facts of human nature and the human situation. Throughout this article, nothing more, or less, will be meant by the autonomy of the tendency of the forces to develop.

Now, the tendency to development here asserted on an asocial basis is supposed, in KMTH, to explain some very important social facts, and, in particular, the supposed fact that

- (3) Social structures have, by and large, been propitious for the development of the productive forces.⁵
- (2) would explain (3) in that, if there is an autonomous tendency for the productive forces to develop, then social relations must be such as to allow it.⁶ It is convenient to give the claim that (2) explains (3) an independent billing:
 - (4) Because there is an autonomous tendency for the productive forces to develop in history, social structures are so shaped or selected as to allow for that development.

Now, which relations of production are propitious at a given stage of development, and, hence, which relations will be selected at that stage, depends on the level of productive power already achieved at that stage. Thus, (4) sustains (5), which was called, in KMTH, the *Primacy Thesis*, since it assigns explanatory primacy to the productive forces over the production relations:

(5) The nature of the production relations of a society is explained by the level of development of its productive forces (134).

⁵ We shall sometimes express (3) by saying, more tersely, that social structures have been (or are) propitious

⁶ For a fuller exposition of the derivation of (3) from (2), see KMTH, 158/9. For a fuller one still, see Andrew Levine and Erik Wright, "Rationality and Class Struggle," New Left Review, No. 123 (Sept./Oct. 1980) Collection, Haridwar CC-0. In Public Domain. Gurukul Kangil Collection, Haridwar

In an impressive critical notice of KMTH, Joshua Cohen⁷ maintained, among other things, that the argument described above is fatally flawed, since asocial premisses could not, he contended, suffice to justify the claim [claim (1)] that the productive forces tend to develop throughout history. If claim (1) is true, then the fact that social structures have been propitious [claim (3)] is an irreducible part of its explanation. It would follow that any tendency for the productive forces to develop is not autonomous, and, consequently, that the primacy thesis is false: a tendency to productive development could not explain why social relations are propitious if their being so is part of the explanation of that very tendency. Joshua Cohen concluded that there is an unavoidable circle in the argument for (5), or a circle to be avoided only by abandoning the attempt to explain (1), and, a fortiori, the attempt to explain it by reference to asocial facts only. But to avoid the circle by adopting the latter course is, so Joshua Cohen claimed, to "enfeeble" the argument for the development thesis (ibid., p. 265).

I.

In this article, we refute Joshua Cohen's claim that KMTH's argument for the development thesis is circular or feeble, after first addressing an earlier critique of the argument made by Andrew Levine and Erik Wright, which to some degree anticipates Joshua Cohen's objections to it. But, before we take up those tasks, more must be said about the nature of the argument for the development thesis.

The argument employs three premisses. The first is that the historical situation of humanity is one of material scarcity: given the character of external nature and the forces available for dealing with it, human beings can satisfy their wants only if most of them spend the better part of their existence engaged in more or less repugnant labor. The second premiss is that people have the intellectual and other capacities needed to discover new resources and to devise productivity-enhancing skills and tools. And the third premiss is that they are rational enough to be able to seize the occasions their capacities create to make inroads against the scarcity under which they labor. In brief: given their rationality and their naturally inclement situation, people will not endlessly forego the opportunity to

⁷ Joshua Cohen, review of KMTH, this JOURNAL, LXXIX, 5 (May 1982): 253–273. ⁸ *Op. cit.*, pp. 61–63.

⁹ KMTH, p. 152. Joshua Cohen calls them the "Smithian premisses," at op. cit., p. 263.

On the present meaning of 'historical', see KMTH, pp. 23, 152. The historical situation of humanity excludes conditions in which nature is bountiful without human assistance, and in which, by virtue of centuries of historical human effort, it has been made bountiful.

expand productive power recurrently presented to them, and productive power will, consequently, tend, if not always continuously,

then at least sporadically, to expand.

Now, KMTH did not say what was meant by "tendency" in the present context, and we shall not give a full answer to that question here. It will suffice to say that the tendency must not be a mere tendency, where that means at least that it must not be regularly unfulfilled because relations of production are unpropitious. II That leaves a lot of reasons why it might, in particular cases, be unfulfilled. Nonfulfillment might be due to material misfortune, such as an earthquake, or a gradual depletion of a needed resource. It might also be due to the scarcity of a particular resource not previously needed, but now needed for further advancement (for example, a certain sort of metal). It might, again, be due to the impossibility, under any production relations, of generating a surplus of dimensions now necessary for progress, because, for example, progress now requires a vast irrigation system; or, more improbably, to a defective physical theory whose errors have not, up to now, created impediments to progress. The relevant common feature of these possible blockers of productive progress is that, in each case, progress is blocked for reasons unconnected with the character of the governing social structure. If such material blockers were very common, then it might be inappropriate to posit a standing tendency of the productive forces to develop, but we conjecture that they are not very common, and no one has criticized KMTH by urging that they might be.

In short, the asocial premisses are intended to establish the existence of a normally fulfilled tendency, one which may be unfulfilled, but which is not regularly unfulfilled for social reasons. And the point of placing this burden on those premisses is that they would otherwise not support a development thesis from which the primacy thesis could be derived. In a different context from the present one, one might speak of a tendency to development which was rooted in subsocial conditions but regularly blocked by social structures, but that is not our usage here, since such a tendency could not explain the character of social structures. ¹²

Note that the claim that there is an autonomous and, by and large, fulfilled, tendency for the productive forces to develop, differs in two important ways from the statement that the forces do, by and large, develop. The first is that they might develop for a miscellany of

11 Cf. Cohen, op. cit., pp. 262/3.

¹² The critics addressed below lacked the benefit of the foregoing partial clarification of 'tendency', but, to the extent that the clarification modifies KMTH, it makes it more vulnerable to their arguments Gurukul Kangri Collection, Haridwar

reasons, and not in fulfillment of any tendency. And the second is that they might tend to develop simply because relations tend to be propitious, and that explanation of their tendency to develop would deprive the tendency of its autonomy.

It bears mentioning that G. A. Cohen was aware that the premisses of the argument for the development thesis were remarkably exiguous, given the fullness of its conclusion. He said that attempting any such argument was "venturesome and perhaps foolhardy." He proceeded nevertheless, because he thought it would be worth finding out what was wrong with an argument of this degree of simplicity and ambition. We shall claim that, whatever else may be wrong with it, it does not suffer from any circularity of the sort its critics discern in it.

II.

According to the argument for the development thesis, human beings are rational and innovative creatures who have a scarcity problem, which they contrive to alleviate by improving their forces of production. A natural, but, as we shall see later, incorrect understanding of the argument sees it as requiring that the agents who actually introduce better forces do so in order that their own burden of labor will be lightened. The picture unintentionally suggested by KMTH is of individual producers, or cooperating groups of them, striving to upgrade their skills and means of production, so that labor will lie less heavily upon them, a picture in which global productive progress is the aggregate result of those several strivings. Following Levine and Wright, be can call this the Rational Adaptive Practices (or RAP) view of the development of the forces, and it is to their criticism of the development thesis, so understood—for that is how they understand it—that we first turn.

Levine and Wright claim that the RAP account of productive development is falsified by the known record of history: in the feudal period, for example, the motivation for such productive development as actually occurred came from the prevailing relations of production, and not from the material situation of producers, considered in abstraction from those relations. The "impulse" to progress, to the extent that there was one, reflected the "class-specific rationality" associated with the position of feudal lords within the prevailing relations of production. From the peasant producer's point of view, "there was nothing 'rational' about the way in which

¹³ Later, the argument is somewhat enriched, through the addition of a subargument which is meant to handle certain objections to it. See KMTH, pp. 153 ff., and the summary on p. 185 below.

¹⁴ KMTH, p. 150. ¹⁵ *Op. cit.*, p. 54.

feudalism allowed for the development of the productive forces" (*ibid.*, p. 63). Development did not occur because of "a rational desire to augment productive capacities in the face of material scarcity, but as an indirect effect of feudal relations" (*ibid.*, p. 62).

Having satisfied themselves that the RAP view of productive development is false, Levine and Wright go on to suggest a remodelled argument for the development thesis which accords with the facts upsetting the RAP account. There might be a tendency to productive progress in virtue of the successive "class-specific rationalities" that arise out of successive sets of social relations. But this fresh case for the development thesis, that is, for (1), would be purchased at the expense of rejecting (2). Productive development would no longer be autonomous, and the attempt to explain material progress in material terms would be abandoned. Moreover, and relatedly, one could now no longer use (1), the development thesis, as a basis for asserting the primacy thesis, i.e., (5). For such a use of the development thesis requires that it explain why relations of production are propitious, that is, why (3) is true, and the revised argument for (1) bases the defense of (1) on an independent assertion of (3). That revised argument is, consequently, a failure, since, as Levine and Wright correctly say,

the explanatory asymmetry Cohen accords the productive forces depends upon an independent argument for the development of the productive forces, one that does not itself hinge on the form of the relations of production.¹⁶

One cannot, without vitiating circularity, defend (1) by means of (3) and also derive (5) from (1). Circularity is avoided if one reverts to the RAP account of the development thesis, in which the impulse to productive progress lies outside the relations of production. But the RAP account is empirically false. The upshot is a dilemma. Either (i) maintain a RAP account of the development thesis, thereby preserving the autonomy of the tendency to development, and enabling a noncircular derivation of the primacy thesis, the entire procedure, however, resting upon an empirically false premiss; or (ii) adopt an empirically defensible "class-specific rationality" interpretation of the development thesis, but then the autonomy of productive development is lost and the attempt to infer the primacy thesis involves a vitiating circle. Asocial premisses cannot explain a tendency to development

¹⁶ Ibid., p. 63. Levine and Wright add that "the development thesis cannot both follow from the primacy thesis and, at the same time, be a presupposition of it." But which is used in the remodelled argument for (1). That remodelled argument bases basis for asserting (5),

(1) on (3), and the problem is that, when (1) is based on (3), (1) cannot be used as a CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

opment which obtains only because class-specific rationalities have the character they do.

We now show that the Levine/Wright critique mislocates the autonomy that the tendency to productive development is meant to possess. Once that autonomy is properly sited, it will be possible to offer a non-RAP account of the development thesis which accommodates the facts upsetting the RAP account, preserves the autonomy of the tendency to productive progress, and enables a noncircular derivation of the primacy thesis.

The key point is that the full development thesis, to wit, that

(2) There is an autonomous tendency for the productive forces to develop.

is different from the claim with which Levine and Wright¹⁷ evidently identify it, namely that

(6) There is a tendency for the productive forces to develop autono-

(6) says that the forces tend to develop without assistance, and, therefore, without the assistance of the relations. But (2) does not entail (6). (2) assigns autonomy to the tendency of the forces to develop, not to the development it is a tendency to.

A child has an autonomous tendency to grow up. He is born with a disposition to do so which is not externally instilled in him by, for example, his parents. But it does not follow that he has a tendency to grow up autonomously, where that means independently of parental and other assistance. The asserted autonomy of the tendency of productive power to grow is relevantly similar. The tendency's explanation lies not within social relations, but in the subsocial facts about humanity rehearsed in section I above. Yet it does not follow that productive power grows without social relational assistance, nor even that social relations cannot be, what they are in Levine and Wright's presentation of the feudal case, and what they certainly are in the case of capitalism, the immediate source of the development of the forces. 18 That the tendency of the forces to develop is realized

on the compatibility between the full development thesis and the fact that capitalist relations induce productive development, see G. A. Cohen, "Reply to Four Critics," Analyse und Kritik, v, 2 (December 1983): 196, 201, parts of which are reproduced in the present section.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

¹⁷ Others, too, lean toward this identification, such as Richard Miller, when he writes that "in a Marxist technological determinism, economic structures . . . have some impact on productive forces (otherwise they could not facilitate their growth) but the most basic changes in the productive forces result from their own expansion in people's pursuit of greater productive efficiency" [Analysing Marx (Princeton, NJ: University Press, 1984), p. 180]. Counterposing the result of that pursuit to the impact of economic structures strongly suggests that the pursuit achieves its result independently of the impact of those structures, and that is the RAP view.

through the specific social relations of particular societies does not contradict the claim that it is rooted in material and, therefore, socially unspecific circumstances of human nature and the human condition.

This clarification of autonomy enables us to present a non-RAP account of the development of the forces, one which, moreover, is congruent with certain utterances of Marx. Consider, for example, this statement, from his letter to P. V. Annenkov of 1846:

. . . in order that they may not be deprived of the result attained, and forfeit the fruits of civilization, [people] are obliged from the moment when their mode of intercourse no longer corresponds to the productive forces acquired, to change all their traditional social forms. 19

Texts like the Annenkov letter confer a Marxian pedigree on the use of human rationality as a basis for asserting the development thesis. But now rationality functions at a different point from that at which it figures on the RAP view, which such texts do not support. The non-RAP claim is not that rational producers introduce superior forces in order to lighten their own labor: that this occurs is not denied, but it is not put forward as the general case. Instead, the claim is that, being rational, people retain and reject relations of production according as the latter do and do not allow productive improvement to continue. In the singularly apt formulation of Philippe Van Parijs,20 the non-RAP claim does not posit a "search-andselection process which operates directly on the . . . productive forces" but "one which operates on the relations of production, which in turn control the search-and-selection of productive forces" (ibid., p. 96). This non-RAP reading of the argument for the development thesis preserves the autonomy of the tendency of the forces to develop. The tendency is not now seen as an effect of the "classspecific rationalities" attached to given sets of social relations; on the contrary. Particular class-specific rationalities prevail only as long as they are associated with class structures that serve a more basically grounded impulsion to productive progress.

In this new use of them, our three asocial premisses (see p. 173 above) generate a (full) development thesis from which the primacy thesis can now, without circularity, be derived. For we can now say that the dilemma formulated above (p. 176) rests upon a false dichotomy. According to Levine and Wright, productive progress under

20 "Marxism's Central Puzzle," in T. Ball and J. Farr, eds., After Marx (New York: Cambridge, 1984).

¹⁹ Karl Marx and Frederick Engels, Selected Correspondence (New York: International Publishers, 1936), p. 8; and see KMTH, pp. 159-160, for comparable statements from other works. (In 1846 'mode of intercourse' occupied the conceptual niche in which Marx later placed '(social) relations of production').

feudalism is either the effect of feudal relations or the result of rationality, intelligence, and scarcity. Yet propitious relations, and their associated class-specific rationalities, might prevail precisely because there is a "universal impulse for human beings to improve their condition by virtue of the kind of animal they are."21 The "rational desire to augment productive capacity in the face of material scarcity" might explain why productivity-enhancing relations obtain (ibid., p. 62). If they indeed obtain, then, by definition, productivity is enhanced because of them, but the sense in which that is true by definition does not deprive the tendency to productive development of its autonomy.22

The KMTH argument may be faulty, and it is certainly not conclusive, but it neither moves in a circle nor denies the truth that relations of production are (at least often) the proximate spring of productive progress. What the argument does deny is that relations of production constitute any part of the ultimate reason why development tends to occur. For it locates that ultimate reason in the facts that people are rational, innovative, and afflicted by scarcity. The conclusion from those facts, that productive forces will tend to develop, would not be true unless social relations were characteristically propitious. But the fact that they are propitious is not why the tendency to development obtains, nor even why, in the final analysis, it is fulfilled, since, in the final analysis, it is because of that autonomous tendency that relations are as they are.

The following analogy might illuminate our reply to the circularity objection. One might argue that there will be a tendency for people's illnesses to be cured by other people, who get better over time at curing given illnesses, simply because people dislike being ill, something which is not, in the present sense, a social fact. Yet suppose, quite plausibly, that, without appropriate medical organization, which is a social structure, little curing, or progress in curing, would occur. The basic explanation of the tendency of illness to be cured, and of its fulfillment, might nevertheless remain, as was claimed, people's hatred of illness, which could also be adduced to explain, through the tendency to improvement in cures which it generates, the existence of propitious medical organization. The route to the latter claim does not seem to involve a circle, and it looks as though, if there is no circularity here, then there is no circularity in the apparently parallel argument for the primacy of the productive forces over the production relations. The relevant general point is that the fact that proposition q is a necessary condition of the truth of

Levine and Wright, op. cit., p. 62. Or, more accurately, by virtue of the kind of animal they are and the fundamental material features of their situation. See, further, KMTH, ch. VI, section (5), especially pp. 160–162.

CC-0. In Public Domain. Gurukul Kangri Collection, Handwar

proposition p does not disqualify the claim that p explains why q is true. That structure involves no circularity in general, so why should one think that a circularity appears when p says that the productive forces tend to develop and q says that social structures are propitious for their development?

Joshua Cohen purports to supply an answer to that question. His answer is that it is impossible to explain the tendency to development, if there is one, without drawing on the proposition that relations of production are propitious. But, if that social fact is an irreducible part of the explanation of material progress, then one cannot, on pain of circularity, invoke a tendency to material progress in

explanation of that very same social fact.

Joshua Cohen might accept that, as we argued against Levine and Wright, the fact that the development thesis is true only if social relations are propitious does not itself preclude an asocial explanation of the development thesis. But he believes that no asocial explanation of it will work: unless one invokes the propitiousness of social relations, one will be unable to explain any tendency to development. (Joshua Cohen's argument for that conclusion is presented in sections III and V below).

Now, Joshua Cohen notes that one might adduce the social fact that relations have allowed development to proceed not in order to explain, but in order to provide further evidence for, the development thesis, beyond the evidence for it which the asocial premisses provide. That is how, he correctly observes, G. A. Cohen used the fact that relations have been propitious in a bolstering subargument for the development thesis, which is summarized on p. 185 below. But, so Joshua Cohen claims, unless one puts that fact to explanatory use, one will not have explained "why there is a tendency to productive progress, and, a fortiori," one will not have explained "this tendency in terms of a few elementary facts about human nature," or, more generally, in purely material terms. He concludes, similarly to Levine and Wright, but on partly different grounds, that the primacy thesis cannot be derived from the development thesis.23 Such a derivation requires that the development thesis be established "on non-social grounds, i.e., without making reference to social forms (especially property relations)." But then the case for the primacy of the productive forces collapses, since there cannot be an "asocially based (autonomous) tendency to productive growth."24

²³ For Levine and Wright, history shows that productive progress comes from class-specific rationality: it is therefore not, as a matter of fact, asocially explained. By contrast, Joshua Cohen thinks a tendency to productive progress could not, as a matter of principle, be asocially explained. His circularity objection does not depend upon the historical record: see sections III and v below. CC-0. In Public Domain: Gurukul Kangri Collection, Haridwar

Joshua Cohen's challenge can be presented as a trilemma. One alternative is to cite no social facts at all in arguing for the development thesis: but then the argument for it is too feeble, as G. A. Cohen recognized—hence his introduction of a subargument, in which social facts are cited. The second alternative, embraced by G. A. Cohen, is to cite social facts as further and indirect evidence for the development thesis, without relying on them to explain it. This may improve the case for believing the development thesis, but now it remains unexplained, since, so Joshua Cohen contends, the asocial premisses are unable, by themselves, to explain it. The third alternative is to accept that social facts must appear in the explanation of the development thesis, but then the primacy thesis cannot be based on the development thesis, without patent circularity. ²⁵

We shall defend the second alternative by arguing that Joshua Cohen has not shown that social facts are needed to explain a tendency to development. The crucial social fact is that "social structures have been such that individual pursuit of material advantage has issued in productive growth." Joshua Cohen holds that we cannot forswear explanatory use of that fact. For if, in particular, we rely for explanation of the tendency to development on the asocial premisses alone, we shall be defeated by the "coordination problem," which is that

the promptings of reason are directed to individuals, whereas the deeds of society depend upon the actions of groups of individuals, diversely prompted by reason (ibid., p. 264).

There is, as G. A. Cohen put it, a "shadow between what reason suggests and what society does," which the asocial premisses cannot remove.

²⁵ We construct this trilemma on the basis of an important but difficult paragraph in Cohen, op. cit., which overlaps pages 264/5. We have interpreted it under Joshua Cohen's guidance. The paragraph is difficult because of an apparent shift in its author's intention which occurs on the ninth line of p. 265. Up until that line, the problem seems to be this: either one says that structures have (merely) as a matter of fact been propitious for productive development, in which case the argument for the development thesis is too feeble, since no tendency to productive growth could be inferred; or one says that it is in the nature of structures that they tend to be propitious, in which case one can indeed infer that the development thesis is true, but one may not then, on pain of circularity, use it to argue for the primacy thesis. After the ninth line, the problem seems, instead, to be this: either one uses the social fact that structures are (or tend to be) propitious to explain the development thesis, in which case circularity results; or one uses it merely as evidence for the development thesis, in which case the latter will lack a convincing explanation. We have opted for the second interpretation, since it is the one Joshua Cohen chose in conversation with G. A. Cohen.

²⁶ Cohen, *op. cit.*, p. 265. KMTH, p. 153.

Now, Joshua Cohen seems to be raising two rather different problems here, each of which is undoubtedly pertinent, but only the first of which should be called a coordination problem.

(i) The first problem is the well-known problem of collective action. Although a given materially possible enhancement of productivity might be in everyone's interest, it might never occur, because it is not in everyone's (and perhaps in no one's) individual interest to initiate, and/or to contribute to, the effort required to bring it about. Although each person would benefit from productive progress, his pay-off structure would be that of a free rider, and so no progress would result. Joshua Cohen grants that this would not always be so, but he argues that "in the abstract," that is, on the basis of asocial premisses alone, there is no reason not to expect "structural arrangements" which "generate undesirable outcomes from individually rational outcomes" to predominate.28 For all that the asocial premisses show, it is just as likely that individuals will be caught in situations without collectively optimal solutions as it is that such solutions will be accessible to them, and there is therefore no general asocially based tendency for development to occur.

This first problem is a genuine coordination problem. Notice, however, that it has nothing to do with the "diverse promptings of reason," on one natural interpretation of that phrase. For in many contexts of collective irrationality, 29 such as, for example, the original Prisoners' Dilemma, everyone is prompted to act in the same direction (because, for example, everyone has the same pay-off orderings), but, unfortunately, they are all prompted away from cooperation. The first problem, then, is that the promptings of reason are directed at individuals rather than at groups, not that individuals (or groups) are diversely prompted by reason.

(ii) The second problem, which is not aptly called a coordination problem, is the problem of (what Levine and Wright called) class-specific rationality. When people are considered in terms of their asocial condition, they indeed appear to have an interest in productive progress, but they may lack that interest, all things (including social

²⁸ Cohen, op. cit., p. 257. The full sentence reads: "For in the abstract there seems no reason to expect 'invisible hands' to predominate over 'prisoner's dilemmas' or other structural arrangements that generate undesirable outcomes from individually rational actions." It would be unfair to place too much weight on this invocation of "invisible hands," but it is important to note that they are by no means the only way that collective action problems are solved or averted. There exist "invisible hands" could not be thought to cover.

²⁹ We use "collective irrationality" merely as a label for the Pareto suboptimal upshot of unsolved collective action problems. We take no stand on the controversial question whether such upshots display genuine irrationality, as opposed to just

conditions) considered, because of the roles they occupy in the social structure. A feudal lord's desire to protect and improve his own material situation would give him an interest *against* productive progress when productive progress would induce changes in social relations which threaten his superior position. The lord's optimal course would then be to oppose improvements in technology, while extracting as much as he could from his peasant dependents. More generally:

. . . the fact that individuals have an *interest* in improving their material situation, and are intelligent enough to devise ways of doing so, does *not* so far provide them with an interest in *improving the forces of production*. Only under *specific structural* conditions is the interest in material advantage tied to an interest in a strategy of productivity-enhancing investment.³⁰

The second problem, then, is that the asocial premisses do not ensure that people's class positions will endow them with an interest in productive progress. It is not the problem that reason prompts individuals rather than groups, for it would obtain even if, for example, reason prompted feudal lords as a class, rather than as individuals. The problem arises because, in virtue of their specific social positions, reason prompts individuals and/or groups diversely, and regardless of the presence or absence of collective action problems within groups.

The first problem is the discrepancy between individual and collective rationality, while the second problem is the discrepancy between what is rational, asocially considered, and what is rational, all things, including class position, considered. But both "discrepant rationality" problems could be thought to generate the same conclusion, which is that, for all that the asocial premisses show, "interests and powers" might be

so organized that progress is blocked, with nothing 'outside' the structure to guarantee its transformation in a way that allows continued progress (*ibid.*, p. 257).

Hence, if there is a tendency to productive development, then not only, and nonfatally, must social structures have a certain shape, but also, and this is the fatal point, the tendency has to be explained (in part) by reference to the fact that they have that shape. The tendency is therefore not autonomous, and it therefore cannot be used to explain the nature of relations of production.

Since Joshua Cohen appears to take the development thesis in a RAP sense (*ibid.*, p. 263), one might respond to the discrepant rationality

Cohen, op. cit., p. 268 CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

problems by exploiting its non-RAP interpretation, as follows: social structures tend to be propitious because, in the last analysis, people do not tolerate them when they are severely unpropitious. And that intolerance has an asocial explanation. It reflects the interest people have, under conditions of scarcity, in progress-inducing relations of production. They are biassed in favor of propitious relations, and that political circumstance is, in the relevant sense, something "'outside' the structure" which tends to ensure that relations continue to be propitious.

But that reply to the discrepant rationality problems is open to an obvious objection. For whether or not the "search-and-selection process which operates on the relations of production" (see p. 178) counts as "'outside' the structure" (by virtue of being 'outside' the economy), that process occurs on a level (within, one might say, a political superstructure) at which both kinds of discrepant rationality problems will once again arise. Thus, Joshua Cohen could now ask: What ensures that political superstructures will enable individuals to organize progressive political change, instead of trapping them in prisoners' dilemmas? And: Why should classes with an interest in enhancing productivity also have the power to replace politically incumbent classes that lack it? If political selection overcomes the discrepant rationality problems at the level of the economic structure, what overcomes those problems at the level of the political superstructure itself? If the asocial premisses do not explain the triumph over rationality problems directly, so that political selection must be invoked, then why suppose that those premisses directly explain the absence of defeating rationality problems in the political superstructure?

Joshua Cohen would conclude that the argument remains "stuck between circularity and enfeeblement" (ibid., p. 265). Our new hypothesis is that, if and when economic structures present insurmountable rationality problems, political superstructures facilitate their replacement. If that hypothesis is intended as an explanatory premiss for the conclusion that the forces tend to develop, then their tendency to do so is not autonomous, and the project of explaining the social by the material collapses into circularity. Recourse to the political superstructure merely expands the circle, for it is part of the social explanandum that superstructures are congenial to propitious economic structures, and that explanandum is here being used to help to explain its material supposed explanans. If, on the other hand, the congeniality of superstructures is, instead, adduced as mere further (nonexplaining) evidence for the development thesis, then the explanation of the development thesis is enfeebled. For asocial premisses cannot explain why rationality problems do not CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar arise at the political level, and they therefore do not explain how they are overcome at the economic level.

Although neither of the discrepant rationality problems described in section III above was expressly formulated in KMTH, the existence of problems of that general kind was noted, and the circularity objection which they threaten to sustain was more or less envisaged (153). Then, to show that, despite such problems, the asocial premisses might strongly support the development thesis, a subargument was offered. Its premiss-called, by Joshua Cohen, the Alleged Facts³¹—was that societies frequently replace productive forces by better ones, and only very infrequently by inferior ones. Its conclusion was that, although there exist obstacles to the expression of rationality, the asocial premisses formulate circumstances that do have a big impact on the world, since, unless one supposes that they do, one cannot explain the contrast between the frequency of productive progress and the rarity of productive regression. Part of the explanation of lack of regression might, of course, be sheer social inertia, but "inertia is too unselective to explain, by itself," a general "lack of regression in face of the fact that there is often conspicuous progress."32

The subargument's conclusion is that the asocial premisses have considerable general explanatory power. That conclusion is intended to suggest that they might have enough explanatory power to justify the claim that there is a tendency to productive progress, which is the conclusion of the main argument. Thus, the frequency of actual progress, in face of the rarity of actual regress, is used to defend the operation of inferring a tendency to productive progress from the asocial premisses. "The facts which best explain why the productive forces frequently advance, but rarely regress, also make it reasonable to believe that the productive forces have a natural propensity to expand."33

Now, Joshua Cohen denies the alleged facts, on which the subargument is based, and we confront that denial in section VI. But he also maintains that the alleged facts do little to overcome the discrepant rationality objection even if they are exactly as alleged. We italicize that clause because we find the thought it expresses extraordinary, and we emphasize that, from now until the end of this section, we shall suppose, with full polemical propriety, that the facts are exactly as alleged.

³¹ Cohen, op. cit., p. 264. ⁹² KMTH, p. 154.

William Shaw, "Historical Materialism and the Development Thesis," Philosophy of the Social Sciences, XVI, 2 (June 1986): 209.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

According to KMTH, the asocial premisses provide a good explanation of the alleged facts, and we may therefore conclude that the conditions those premisses formulate powerfully constrain the structure of society. Joshua Cohen disputes the starting point of that argument. He thinks that, whatever would be the real explanation of the alleged facts, the asocial premisses simply could not explain them adequately. "What the facts, if they were as alleged, would show is that structures have been such that individual pursuit of material advantage has issued in productive growth." They would show that self-interested rationality regularly finds a way through the reefs of suboptimality and counterfinality. But rationality can do that only if there is a way to be found, and the asocial premisses cannot explain why such a way is available. The conditions they formulate are powerless to prevent either of the discrepant rationality problems from being insoluble.

Take, first, the collective action problem. We could imagine a society in which a universal interest in progress was not satisfied, because people were locked into an unmodifiable multiperson prisoners' dilemma. Or, to take the class-specific rationality problem, imagine a society that meets the following specifications. Eighty per cent of its members live off the labor of the rest. Productive progress would so upset social relations as to render that exploiting majority much worse off. And the producers, who have an interest in productive progress, and the technical capacity to bring it about, are politically unable to challenge the majority class (cf. *ibid.*, pp. 268/9).

No matter how rational, intelligent, and motivated to pursue material advantage individuals are, no tendency to productive progress would ensue in these logically possible societies. But, for all that any asocial premisses show, all historical societies could have been relevantly like one or other of them. Hence, no asocial premisses can yield the development thesis. To derive the development thesis, one needs the asocially inexplicable premiss that social structures have as a matter of fact contained channels along which rationality could flow. To explain a tendency to development, that would be required as an independent premiss. And that restores the circularity-or-enfeeblement dilemma.

G. A. Cohen's subargument was offered in mitigation of "two large gaps" in the original argument for the development thesis. The first was that individuals might not care enough about their material problems, since they have other problems to contend with too. The second was the one already mentioned, that "it is not evident that societies are disposed to bring about what rationality would lead men

³⁴ Cohen, op. cit., p. 265.

to choose."35 Joshua Cohen allows that the subargument is a "plausible response to the first difficulty." But he denies that it helps with the second. For "it is not facts about individuals that are now in question, but facts about the structures that determine the outcomes of individual acts."36 The circularity-or-enfeeblement dilemma is unavoidable, since asocial constraints could never ensure that structures will offer opportunities for overcoming discrepant rationality problems. If the facts are indeed as alleged, then that must be because social structures have, on the whole, been favorable to development, and not in virtue of any extra-social developmental tendency.

The above objection does not challenge the alleged facts. It therefore does not say that societies of the logically possible sort it describes occur universally, for it follows from the alleged facts that propitious social relations are very common. What the objection says is that their being common could not be materially explained.

We reply that no good argument has been provided for that conclusion. For one way in which the asocial conditions (whose weight the subargument seeks to enhance) might assert themselves is precisely by militating against widespread occurrence of logically possible societies of an unpropitious kind. Though logically possible, they might, for material reasons, be contingently unlikely, and then the above objection would fail.

How might the facts recorded in the asocial premisses make obstructive structures materially improbable? Those premisses concern not only the motivations and capacities of individuals, but also the fundamental material features of their situation. One such feature is the limited amount of productive power available to them, and that feature is relevant to both of the examples described above.

Thus, to address the first example, prisoners' dilemmas are inescapable only when people are unable to modify each other's pay-off structures in certain ways. But even when they cannot do so, the result might not be so dire, because of the uncertainties and interdependencies imposed by material scarcity. Interdependency tends to make prisoners' dilemmas occur in indefinitely long series, and it is by now well-known that such seriality facilitates their solution. Material conditions might also enable "political entrepreneurs" to modify pay-off structures so that the free rider quandary is transcended.

Consider, now, the second example cited above. A society in which a bare twenty per cent of the population does all the work is, by definition, at a very high level of productive power. But that would

³⁵ KMTH, p. 153.

³⁶ Cohen, op. cit., pp. 264/5.

have liberating consequences for the political capacity of the exploited class, so that the features characterizing the society, while logically compatible, might never co-occur, for material reasons.

Structures do organize "interests and powers," but they do not generate them ex nihilo, and the material constraints under which historical structures arise might regularly open ways for individuals to ensure that productive progress occurs. That seems to be a possible conclusion from the alleged facts. It is not, perhaps, a mandatory one, but it is not in principle excluded in the fashion argued above.

Despite the fact that economic and political structures are not unproblematically congenial to progress, the alleged facts entail that progress wins through. The subargument then presents this challenge: either accept that the asocial premisses have great explanatory power, or offer an alternative explanation of the contrast between frequency of progress and infrequency of regress. Note that the greater the propensity of social structures to throw up rationality problems is, the better grounded is the conclusion of the subargument, if, as we are here allowed to suppose, the alleged facts obtain. Despite the potential for suboptimality and counterfinality in social structures, the facts say that progress often occurs and regress hardly ever does, and that suggests that the asocial conditions of human nature and circumstance are strongly efficacious, for they seem to have the strong effect of subverting persistingly unprogressive structures in favor of progressive ones.

We conclude that the subargument's challenge is intact. Nothing Joshua Cohen says succeeds in dissipating its force.

Here is a further illustration of the importance of material constraints on structures.

Joshua Cohen remarks that the alleged facts are supposed to answer G. A. Cohen's first objection (see p. 186 above) to his argument for the development thesis by "providing indirect evidence that material interests are in fact of paramount significance for individuals" (ibid., p. 264). But the alleged facts might sustain an answer to that objection which takes an interestingly different form,38 because it does not imply anything "definite about the priorities among" people's "different goals." Material constraints might ensure that people's nonmaterial interests, though not at all secondary, tend not to

³⁷ *Ibid.*, p. 257: see above, page 183.

³⁸ To be fair, we note that KMTH provided no hint of this different form of

³⁹ Allen Wood, Karl Marx (London: Routledge & Kegan Paul, 1981), p. 30.

conflict with their material interests in progress-defeating ways. It might even be that their other goals are "preferred to productive development whenever they conflict with it, but" that they "conflict with this goal so infrequently," or at such noncritical junctures, "that they never endanger a materialist account of history" (*ibid.*, p. 30). That conjecture might, of course, be empirically implausible, but the way to show that is to assess the fidelity of the materialist account to the facts of history. The claim that there is an inevitable circularity in any attempt to argue for the account cannot be sustained.

Now, Joshua Cohen does, also, construct a factual case against historical materialism, when he addresses himself to the subargument's claims about the historical record. He does not, of course, question the alleged fact that "productive forces are frequently replaced by better ones," but he denies the subargument's other allegation, which is that "societies rarely replace a given set of productive forces by an inferior one," that there is, in other words, "little productive regression" in history.⁴⁰

Joshua Cohen counterclaims "that there is *substantial* regression and stagnation, alongside productive progress" (*ibid.*, p. 266). Now, (prolonged) stagnation is, in a general way, *prima facie* evidence against historical materialism, but it is not evidence against either of the alleged facts, and here, in an admittedly somewhat narrow way, we shall stick to our brief and examine the adequacy of Joshua Cohen's case for the claim that the record shows "*substantial* regression."

Joshua Cohen's first example is the "long period" of Chinese history "roughly corresponding to the Ming and Ch'ing dynasties," in which, he says, there was "little evidence of . . . development of productive forces, though . . . no apparent regression either" (*ibid.*, p. 267). The latter concession makes the example irrelevant here: what we have is, at most, an example of stagnation. And it is not clear that we have even that much. For Joshua Cohen also says that, in the given period, "agricultural output continues to expand on the basis of the extension of cultivation and increased output per acre on increasingly fragmented holdings" (*ibid.*) and the exploitation of new resources implied by the extension of cultivation could count as an expansion of productive power. Any increase in output per person, or, more accurately, in potential output per person, qualifies, for nonarbitrary reasons, 41 as a development of the productive forces, whether or not the increase is based on an improvement in

⁴⁰ Cohen, *op. cit.*, p. 264. See KMTH, pp. 60/1.

productive technique. Joshua Cohen thinks the China example is a case of stagnation only because he misidentifies the development of the productive forces with the development of technology alone.

In the second paragraph of his historical review, ⁴² Joshua Cohen mentions more cases of putative stagnation, and only one case of a putatively "genuine regression in agricultural productivity, resulting from the 'second serfdom', i.e., from the imposition of substantial labour services on the Polish peasantry" (*ibid.*, p. 267).

Now, the "second serfdom" was certainly a regression in relations of production, since it replaced relatively free by relatively bound labor. One can, moreover, agree with Joshua Cohen that the "imposition of services guaranteed to the Polish lords what was certainly more important to them than an increase in the forces of production, viz., control over the surplus produced in agriculture' (ibid.). But three considerations undermine Joshua Cohen's hope that the example will serve his polemical purposes: (i) The fact that control over the surplus was more important to the lords does not mean that they would not also have cared how big that surplus was, and, therefore, how effective the agricultural productive forces were. (ii) What needs to be shown, to establish regression in the level of development of the forces, is that potential output per person dropped, because, e.g., of a loss of agricultural skill, or a destruction of good means of production. The Polish nobles might have accepted a trade-off in which existing forces were underused, or badly used, but that would not mean that the power of the stock of forces available to them was diminished. (iii) If the revived serf relations later crumbled because the forces they employed were actually and/or potentially underproductive, then that would support the development thesis in an indirect way: it would be evidence that forms of society which do not share in the tendency of the forces to develop are likely to go under.

These brief rejoinders do not, of course, settle the historical issue in our favor. What they do show is that it is more difficult to refute our case than one might think when crucial distinctions are overlooked: between stagnation and regression, between actual and potential output per person, and between technical progress and productive progress in a comprehensive sense. Not being historians, all we can offer are these potent conceptual reminders—and one important nonconceptual point:

⁴² Cohen, op. cit., pp. 267/8.

In light of the fact that it is only relatively recently that historians have come to appreciate the full extent of technical progress both under feudalism and within the ancient world—both of which had long been thought innocent of technical innovation—one needs to avoid facilely writing off earlier societies as devoid of productive progress.43

Underlying Joshua Cohen's criticisms is, he says, a "skepticism" about KMTH's "ambition,"44 which is "to defend a non-theological theory of history in which progress is the central tendency" (ibid., p. 257). Such a defense, he rightly says, requires "some extra-social factor . . . controlling historical change" (ibid., p. 273).

In a domain as uncertain as the present one, everybody is entitled to be skeptical: no one could claim to know that extra-social features of human nature and the human situation operate powerfully enough to generate a historical tendency capable of overcoming recalcitrant social structures. But material conditions patently constrain social structures to some degree, and our chief complaint is against an a priori denial that the constraints are strong enough to induce the mooted extra-social tendency. Joshua Cohen's circularity objection (though not his objection to the alleged facts) amounts to such an a priori denial.

We have acknowledged that social structures must be of a certain shape if there is to be a tendency for the forces of production to develop. But we have denied that that is a reason for rejecting the extra-social explanations of historical materialism. If it were such a reason, then there would be (equally a priori) reasons for ruling out such contributions to social understanding as Marvin Harris's Cultural Materialism, 45 and the whole field of human sociobiology. 46

G. A. COHEN

All Souls College, Oxford University

WILL KYMLICKA

Princeton University

⁴³ Shaw, op. cit., p. 205. For an excellent account of precapitalist productive progress and the mechanisms driving it, see Gunnar Persson, Pre-Industrial Economic Growth in Europe (Oxford: Basil Blackwell, forthcoming). 44 Cohen, op. cit., pp. 254/5.

⁴⁵ New York: Random House, 1979, especially ch. IV.

Whatever may be the defects of sociobiology, it is surely not to be dismissed on the mere ground that complaisant social structures are required to realize the

MORAL JUSTIFICATION AND FREEDOM*

N Elbow Room Daniel Dennett counsels us to consider the varieties of free will worth wanting. His message for philosophers in the compatibilism debate is: Do not worry about whether a variety of free will is compatible with determinism if it is not worth wanting in the first place. I take this to be sound advice. So which varieties of free will are worth wanting? The capacity to act rationally is generally regarded as an important kind of freedom. We think that, when one's capacity to act rationally is diminished, whether because of cognitive or emotive defects, so is one's capacity to exercise free choice; and we believe that rationality in action is worth wanting. But this response to Dennett's counsel only pushes his question back a stage. As with free will, there are different conceptions of rationality. Which of these matter, and why?

At various points in his book, Dennett relies on a conception of rational choice which has become standard in economics and the social sciences generally. It is rationality as efficiency or, more technically, individual utility maximization. There are, of course, subvarieties within this standard. The term 'individual utility maximization' covers a family of formulations that are not extensionally equivalent. But its members are similar enough in content and origin to be a family. On this conception of rational choice it is obvious why rationality matters. To act rationally just is to promote as much as possible what one believes is worth wanting. It is, in a variant terminology, to maximize the satisfaction of one's considered preferences. No wonder, then, that the variety of free choice that is identified with rational choice should be worth wanting. Not that one's beliefs about what is worth wanting are infallible. The point is rather that, since one's considered preferences constitute significant evidence for what is worth wanting-perhaps the best evidence that one can have at any given time—this kind of freedom, we can call it the Freedom of Instrumental Rationality, is inherently and self-evidently worth wanting.

The freedom of instrumental rationality, however, may seem to be tangential to the traditional problem of free will. Since the seven-

Elbow Room: The Varieties of Free Will Worth Wanting (Cambridge, MA:

MIT Press, 1985).

0022-362X/88/8504/0192\$02.20

© 1988 The Journal of Philosophy, Inc.

^{*} Earlier versions were presented at Dalhousie University and McGill University. I am grateful for Susan Sherwin's help with Part IV and for the written comments received from Neera Badhwar, David Braybrooke, Robert Bright, S. A. M. Burns, David Copp, David Gauthier, Duncan MacIntosh, Ray Martin, Ann Wilbur Mac-Kenzie, J. Howard Sobel, Terry Tomkow, and Kadri Vihvelin.

teenth century, the central issue has been whether determinism is compatible with moral agency. From this perspective, the variety of free will worth wanting is the freedom of choice essential to being a moral agent. Such freedom is not obviously reducible to the freedom of instrumental rationality. Therefore, even if one could make a good case for the compatibility of determinism and instrumental rationality, the traditional dilemma would appear to remain. But recent developments in moral philosophy suggest that this appearance is misleading. An increasing number of writers2 have views that approximate the following thesis: To justify a system of moral norms for society is to demonstrate that, if each of its members were to choose only under the constraint of instrumental rationality, each would choose to have those norms govern their behavior. If this thesis is basically correct, then questions about the limits of moral justification and moral agency, including the traditional question of whether moral freedom is compatible with determinism, would be reducible to questions about the limits of instrumental rationality.

The un-Kantian assimilation of freedom and morality to instrumental rationality is not new in Western philosophy. Thomas Hobbes made it in his social-contract theory of justice and so, arguably, did David Hume in his. What is new is the uncompromising character of the contemporary assimilation, coupled with a precise interpretation of instrumental rationality as individual utility maximization defined in the language of Bayesian decision theory and the mathematical theory of games. Since the concepts of these theories seem, however normative, to be without any specific moral content, the assimilation has the character of a reduction. This point deserves emphasis. John Rawls and John Harsanyi explicitly use the concepts of decision and game theory in constructing moral theories as part of the theory of rational behavior.³ But neither theorist attempts a reductive account of morals. Compare the standpoint of Gauthier stated in the opening pages of Morals by Agreement:

To choose rationally, one must choose morally. This is a strong claim. Morality, we shall argue, can be generated as a rational constraint from

See John Rawls, A Theory of Justice (Cambridge, MA: Harvard, 1971); and John Harsanyi, "Morality and the Theory of Rational Behavior," in Amartya Sen and Bernard Williams, eds., Utilitarianism and Beyond (New York: Cambridge,

² See, for example: Richard Brandt, A Theory of the Good and the Right (New York: Oxford, 1979); James M. Buchanan, The Limits of Liberty: Between Anarchy and Leviathan (Chicago: University Press, 1975); David Gauthier, Morals by Agreement (New York: Oxford, 1986); and Jan Narveson, "Remarks on the Foundations of Morality," read at the Canadian Philosophy Association meeting in Win-

the non-moral premises of rational choice. Neither Rawls nor Harsanyi makes such a claim. Neither Rawls nor Harsanyi treats moral principles as a subset of rational principles for choice (4).

The view that moral principles are a subset of instrumentally rational principles for choice I shall label the *Reduction Thesis*.

The reduction thesis has two implications that are relevant for the present discussion.

(RT1) Any variety of freedom compatible with being a moral agent must be compatible with an agent's being instrumentally rational.

(RT2) Any variety of freedom that is inessential to an agent's being instrumentally rational is inessential to his being a moral agent.

These implications hold on the assumption, which I shall not defend here, that a moral agent is an agent who has the capacity to act on moral principles and an instrumentally rational agent is an agent who has the capacity to act on instrumentally rational principles, including any which are also moral principles. My main aim in this paper is to explore the bearing of the reduction thesis on the freedom of moral agents. In the first part, I shall argue on the basis of (RT1) that the reduction thesis is inconsistent with a certain libertarian (hence incompatibilist) conception of free will4 when the morally relevant principles of choice are as Gauthier understands them to be. In the second part, I shall respond to several objections which will allow me to locate the topic of the first part within the context of larger issues concerning free choice. In the third part, I shall argue on the basis of (RT2) that there is good reason to believe that the reduction thesis is incompatible with libertarian free will independently of the details of Gauthier's analysis of rational choice. Finally, I shall argue in the fourth part that there are limitations on the reduction thesis which (1) do not undermine the importance of the inconsistency argued earlier but (2) reveal obstacles to free choice masked by presuppositions on both sides of the compatibilism debate.

I.

I begin then with the conception of moral justification advanced by Gauthier in *Morals by Agreement*. His view is that behavior is morally justified when it conforms to social norms which are themselves morally justified. These are the norms which could and should be embodied in the social institutions of a society. How are these norms of behavior themselves morally justified? The answer to this question

⁴ By 'libertarian' I shall mean someone who holds that people sometimes exercise free will and that this exercise of free will is incompatible with determinism. For the purposes of this paper, it will not be necessary to settle on a definition of 'determinism'.

contains the reduction. Such norms are morally justified when, but only when, the members of society would agree, as instrumentally rational agents, to have these norms constrain everyone's behavior and each would be motivated, as an instrumentally rational agent, to comply with these constraints when the rest do. Notice that this reductive conception of moral justification implies two conditions for success. Instrumentally rational agents must be able to reach mutual agreement about the system of social constraints without compromising their rationality and each must be rationally motivated to comply with the constraints provided that the others do.

Corresponding to these conditions are two major problems that Gauthier addresses in his book (in chs. 5 and 6, respectively). The first he calls the Bargaining Problem and the second the Compliance Problem. The bargaining problem is to state in a general and precise way what the agreement among instrumentally rational agents should look like. One requirement for the agreement seems obvious. The pattern of social interaction agreed upon should be Pareto-optimal. That is, there should be no alternative outcome of individual actions which would be better for at least one person and no worse for the others. For if there were such an outcome, then at least one person would have an instrumentally rational basis for rejecting the proposed agreement in favor of this alternative, and no other person would have an instrumentally rational basis for not going along with the first. But Pareto-optimality is in general only a necessary condition for a solution to the bargaining problem. More than one optimal outcome may be possible, and each of them may be very unsatisfactory to at least one person. In this kind of conflict-ofinterest situation, the solution, if it is possible at all, is surely nontrivial.

I shall not present Gauthier's ingenious general solution, since it is the next part of his theory of moral justification which bears directly on the compatibilism debate. This part concerns the compliance problem, which arises because an optimal outcome that fits the solution to the bargaining problem cannot be expected to be a Nash equilibrium in every case. A Nash Equilibrium outcome is a set of actions, one for each person in a social interaction, such that, for each person, there is no alternative outcome that this person would prefer given the actions of all the others. Clearly, if a set of independent actions selected by the solution to the bargaining problem is not a Nash equilibrium, then some person will have an instrumentally rational basis for not doing his part in producing the agreed upon outcome. We might wish to say that such noncompliance on his part, though instrumentally rational, is not morally justified. But to give

this response would be tantamount to giving up the reduction thesis. If the reduction thesis holds, each person must, as an instrumentally rational person, be motivated to comply with the constraints imposed by the solution to the bargaining problem. But how is this possible? This is the compliance problem.

The logic of the compliance problem is nicely illustrated by the notorious Prisoner's Dilemma (PD). Saving complications for later, let us focus on the simplest case: the two-person, two-action, one time only PD, in which the agents' actions are mutually inde-

pendent.

Column

		Cooperate	Defect
Row	Cooperate	B,B	D,A
	Defect	A,D	C,C

The letters in the outcome matrix show the preferences of row and column, A for most preferred outcome, B for next most preferred, and so on, with row's preferences shown to the left of column's. There is one Nash equilibrium point: mutual defection (C,C), which is suboptimal. The other outcomes are all Pareto-optimal. In this special case, the solution to the bargaining problem is trivial. Only mutual cooperation (B,B) is an optimal outcome which both row and column can agree to support from the standpoint of instrumental rationality. The compliance problem, on the other hand, appears to be utterly insoluble. Instrumentally rational agents, who will never meet again and who cannot causally influence each other's choices, appear bound to defect. Each has a (strongly) dominant strategy: defection. Under the condition of causal independence, defection results in gain no matter what the choice of the other party. In a phrase, defection maximizes individual utility.

I take the apparent insolubility of the compliance problem to be its relevance for the problem of free will. Let us imagine the PD from the perspective of a libertarian. To give the example some poignancy, imagine also that each player in this game believes that the other will cooperate. (This assumption makes us depart slightly from the classic conception of the game, but eventually, for the sake of realism, we will have to anyway. The assumption certainly does not ease the compliance problem.) Each player believes that, if he cooperates, he gets his B outcome, and, if he defects, he gets his A outcome. Of course he is tempted to defect. At the same time, we may suppose, he feels a great reluctance to defect. Not that he does not truly prefer the A outcome to the B outcome. Indeed, he does! It

is his moral training which gives him pause. He has been brought up to think that defection against a cooperator is wrong. From the libertarian perspective, what will he do?

We must be careful here. I do *not* mean that, when guilt is added in, the expected A outcome is no longer preferred to the B; if that were so, the pay-off structure would no longer exemplify the PD. No, even considering the expected guilt, each prefers outcome A to outcome B. Moreover, each can resist any moral or sympathetic urge to cooperate. From the perspective of the libertarian, any agent can. That is, in circumstances contemplated, where one is tempted by advantage to act immorally, one can resist the pressures created by moral training or sympathy or any other kind of psychological propensity, and choose between the alternatives by an exercise of free will. On what basis then will the decision be made?

Let us assume that each agent is fully rational, i.e., each will act in accord with the principle that a person should maximize expected individual utility; then it follows that the agents will defect. This follows from the reduction thesis and the libertarian perspective on choice. Under these assumptions, there is no general solution to the compliance problem. Instrumentally rational agents will defect in this PD, if they have the ultimate power to act independently of all psychological constraints other than those which define relevant parameters of the decision problem.

So much the worse for the reduction thesis, the libertarian might say. But, before we side with the libertarian, we should consider how Gauthier deals with the compliance problem. Although he does not say that his solution favors compatibilism (he does not mention this debate in his book), I shall argue that his solution is incompatible with the libertarian conception of moral agency.

On Gauthier's view, the fundamental unit of moral evaluation is not the agent's particular choice, such as cooperation or defection in the PD, but rather the disposition to choose which the particular choice expresses. The following words are from his chapter on compliance:

⁵ Is this fair? Peter van Inwagen, who defends an incompatibilist conception of free will, appears to be committed to this implication in *An Essay on Free Will* (New York: Oxford, 1983), p. 63f; but, in a recent paper ("When Is the Will Free?" delivered at a conference on Freedom and Mind in Montreal, September 1986), he has argued that the will is not always or even usually free. It is important to note, however, that my argument does not assume that, on a libertarian conception of free will, the will must always or usually be free. My argument requires only that the will be free on this conception in a situation of the kind described, for example, one in which a moral agent is aware of having to make a moral decision and finds himself tempted to do otherwise than what he has been taught is morally right. I assume that any interesting libertarian conception of free will is committed to this much.

The received interpretation, commonly accepted by economists and elaborated in Bayesian decision theory and the von Neumann-Morgenstern theory of games, identifies rationality with utility-maximization at the level of particular choices. A choice is rational if and only if it maximizes the actor's expected utility. We identify rationality with utility-maximization at the level of dispositions to choose. A disposition is rational if and only if an actor holding it can expect his choices to yield no less utility than the choices he would make were he to hold any alternative disposition. We shall consider whether particular choices are rational if and only if they express a rational disposition to choose (182/3).

Two dispositions to choose are salient. Gauthier calls one Straightforward Maximization (SM). This disposition to choose is extensionally equivalent to choosing rationally according to the received interpretation of rationality, that is, utility maximization at the level of particular choices. The SM (person) will simply choose whichever course of action would maximize his expected utility at the particular time of choice. The SM would therefore defect in the PD. The other disposition, called Constrained Maximization (CM), embodies a conditional strategy with respect to the PD. If the CM believes that the other party will cooperate, then the CM will too; otherwise not.⁶ The intuitive idea is that the CM aims at that optimal outcome that is selected by a solution to the bargaining problem. Thus, in the PD, the CM aims at mutual cooperation. Failing any expectation of success in reaching this outcome, the CM will revert to behaving like the SM. The upshot is that, when a CM meets a SM in a PD and recognizes the SM to be such, there will be mutual defection just as when SMs meet; but, when CMs meet each other in the PD and recognize each other to be such, there will be mutual cooperation.

We can now summarize Gauthier's solution to the compliance problem and show its incompatibility with the libertarian perspective. To solve the compliance problem in the PD, Gauthier must show that it is rational for someone to cooperate if she believes the other party will in fact be cooperating on the assumption that she will too. Now, it is rational for her to cooperate if, and only if, her cooperation would express a disposition to choose which is itself

⁶ I have deliberately oversimplified Gauthier's formulation (167–170) in the interest of avoiding complications that do not pertain to the case at hand. Even in this case, however, one can worry about the possibility of an impasse, for each CM is prepared to cooperate if, and only if, the other will. How does either know the other will cooperate? It appears as if an implicit circularity (or regress) in this formulation of the disposition threatens to immobilize the actors. I propose a solution to this problem in my critical notice, "Gauthier's Theory of Morals by Agreement," *Philosophical Quarterly* (forthcoming).

rational. This is Gauthier's assumption, expressed in the passage quoted, that the rationality of a particular choice depends on the rationality of the disposition to choose which the particular choice expresses. Call this his *Dependency Thesis*. The disposition to choose which would yield cooperation in the circumstances described is the CM disposition. The compliance question thus becomes the question whether the CM disposition is rational. That disposition is rational, of course, just in case a person with that disposition can expect to maximize utility.

We have, therefore, a second-order decision problem to solve, where the salient choices for each party are the two dispositions: CM and SM. The second-order problem can be a decision under risk in which the beliefs about the other person's disposition have only a probability less than one of being true (see Gauthier, pp. 174–177). Since that complication will not alter the logic of the case against libertarianism, I shall stick to the simplest case where the dispositions of each person will be transparent to the other when they meet in a PD.⁷ The second-order choice matrix looks like this:

		Column		
Row	CM	CM Mutual Cooperation	SM Mutual Defection	
	SM	Mutual Defection	Mutual Defection	

Notice that here there is no second-order compliance problem, for the unique Pareto-optimal outcome, mutual cooperation in the PD, is also the unique equilibrium outcome. Row and column have (weakly) dominant pure strategies, CM in each case; choosing independently they would, as individual utility maximizers, choose CM.

⁷ There is a complication for the case of transparent dispositions which deserves mentioning. We are supposing that the agent's own disposition is transparent to herself as well as to the other agent. Suppose it were not self-transparent. Then, at the moment of choice, each agent would be in a Newcomb problem of the infallible predictor variety. For example, row, knowing column is a CM and hence knowing that column's choice is an infallible prediction of her own choice, but not knowing her own disposition or choice, would know that the conditional probability of column's cooperation, given her own, is one, and that the conditional probability of column's defection, given her own, is also one. Yet her choice has no causal influence on column's and she has a dominant strategy: defection. It could appear, then, that Gauthier's argument that cooperation is rational opposite a known CM in the PD is equivalent to an argument that choosing the opaque box alone is rational in an infallible predictor Newcomb problem. In my critical notice, op. cit., I argue that Newcomb's problem does not arise for Gauthier given his dependency thesis (see above), and the fact that there is no apparent paradox regarding maximization at the level of dispositions to choose.

CM is therefore the rational disposition, and cooperation in the PD, given the dependency thesis, is the rational first-order choice.

Now, we might ask: Is it possible for persons who recognize themselves to be in a PD and to have the CM disposition to override this disposition at the moment of choice? The answer, defended in detail below, is that it is not possible. The CM disposition is so understood that an agent who can (in the libertarian sense) resist, at the moment of choice, any psychological disposition which he has cannot (logi-

cally) embody the CM disposition.

If this is so, it should be clear that Gauthier's solution to the compliance problem cannot be reconciled with the libertarian perspective on moral agency. Under the reduction thesis, in particular under (RT1), the *rational* libertarian agent would behave exactly as if he were bound by the SM disposition. For such an agent, being an individual utility maximizer and able, at the moment of choice, to resist any disposition which he has, will in effect embody the SM disposition and be applying the maximizing standard at the level of particular choices. The result is a collapse of the second-order decision problem into the first and defection in every PD. I see no obstacle to generalizing this conclusion to every choice situation with a suboptimal equilibrium point. Under reduction, no Gauthier-style solution to the compliance problem is available to the libertarian.

The libertarian can, of course, challenge the reduction thesis, but I shall not pursue this option. I am content to let my criticism of libertarianism be conditional on the adequacy of this thesis for at least some range of human choice. My own suggestions about how to supplement it, which are stated below in the fourth part, will be no consolation to the libertarian. In the next part, I shall anticipate three objections that can be answered if the reduction thesis or a restricted version of it is granted.

II.

The first objection can stated quite simply. If Gauthier is right that it is rational for an individual with a CM disposition to cooperate in a PD with someone known to be similarly disposed, then, in this situation, a libertarian free agent can act rationally just by behaving in the same way that somebody with the CM disposition would, namely, by cooperating. If cooperation really is rational for an agent in these circumstances, then a rational libertarian free agent would, of course, cooperate. Although such an agent can do otherwise, being rational, he would not; he would choose to cooperate.

Since this objection is so natural and raises such fundamental questions regarding the proper interpretation of Gauthier's argument, it is best to proceed slowly. First, I want to make clear why I

l lieve that a libertarian free agent cannot embody the CM disposition. Here I shall assume that the dependency thesis holds (i.e., that the rationality of choices depends on the rationality of the disposition that they express). Then, I shall take up the dependency thesis directly.

To make clear why a libertarian free agent cannot embody the CM disposition, we need to return to Gauthier's second-order decision problem and ask how we should complete the upper left cell of the decision matrix.

		Column		
		СМ	SM	
	CM	?	Mutual Defection	
Row	AUG DE LES			
	SM	Mutual Defection	Mutual Defection	

In this decision problem, it will be recalled, we are assuming that row and column are utility maximizers faced with a hypothetical choice between two dispositions. Suppose for the moment that a CM disposition is not causally binding on choice, that is, that an agent can (in some contra-causal sense) act contrary to the disposition at the moment of choice in the PD. How then is the matrix to be completed? One might be tempted to answer that the disposition is *rationally* binding, even if not causally binding, and hence that row and column can expect that they will cooperate if they have chosen the CM disposition. But that answer is question-begging in the present context. For the point of the second-order decision problem is to determine which disposition is the rational disposition to have, i.e., which disposition is utility–maximizing. To assume at this point that rational agents will stick to the CM disposition at the moment of choice because it is rational is to argue in a circle.

Unless all the cells of the matrix are to be mutual defection, we need to understand the CM disposition to guarantee cooperation with a known cooperator, independently of any assumption about the rationality of the disposition. ('Guarantee' is too strong. An indetermininistic disposition, sufficiently weighted toward cooperation with apparent cooperators, is enough. In other words, one could construe dispositions as strong tendencies to choose, provided that the agent cannot alter the tendency at the moment of choice. The second-order decision problem would then be a decision under known risk. I ignore this complication, which is no different in kind

from the one mentioned earlier which Gauthier considers in relaxing the assumption that the agents' dispositions are transparent to each other.) By taking the CM disposition to be one that agent cannot opt out of at the moment of first-order choice and hence to be incompatible with libertarian free agency, we can assure ourselves, without circular argument, that known CMs will operate in the PD.

The next objection is just to challenge Gauthier's dependency thesis. Gauthier assumes that the rationality of a choice depends on the rationality of the disposition expressed by the choice. But why not vice versa? So far no explanation has been given. If the dependency were reversed, Gauthier would have no more solution to the compliance problem than the libertarian has, given that moral agency is instrumentally rational agency.

Gauthier's position contains an answer which goes to the heart of his conception of rationality. It can be put simply. His answer is that the ultimate standard of rationality, the maximization of individual utility, should be applied first at the level of dispositions, because otherwise actions expressing the SM disposition would be rational. We would, therefore, regard a person as being rational at both levels who expects to gain less utility than someone who is rational at neither. For example, if the ultimate standard is applied at the level of choice first, then, in the PD, defection would be rational even if the other party is believed to be a cooperator. Hence, on the reverse of the dependency thesis, the SM disposition would be rational. But SMs can expect to gain less utility on average than CMs. In the language of Gauthier's earlier work,8 the reversal of the dependency thesis makes the resulting conception of instrumental rationality fail to be "self-supporting." Such a conception would be subject to rejection by its own internal standard.

It is possible to put this reasoning in the form of a reductio. Note that having the SM disposition is equivalent to applying the maximizing standard at the level of choices, and having the CM disposition is equivalent to applying the maximizing standard at the level of dispositions to choose. Now, consider the choice between applying the standard at one level and applying it at the other, and suppose that it is rational to apply the standard at the level of choices, so that it is rational to choose to apply it at this level in this choice problem. By the equivalences just mentioned, this choice problem is equal to a choice between having the SM disposition and having the CM disposition. We have already seen that, in this choice problem, applying

⁸ "Reason and Maximization," Canadian Journal of Philosophy, IV (1975):

the maximizing standard yields the conclusion that the choice of having the SM disposition is not rational. But this conclusion entails, by one of the equivalences mentioned above, that it is *not* rational to choose to apply the standard at the level of choices. Contradiction.

These replies to the first two objections lead directly to a third. According to the argument thus far, if Gauthier's solution to the compliance problem is to work, the dispositions to choose must be thoroughly effective in determining choice. That is, Gauthier's maximizing agent, unlike the libertarian free agent, cannot step in and out of dispositions at will. In short, there has to be a relevant contra-causal sense of 'can' in which the CM who believes the other will cooperate cannot do otherwise than cooperate herself. But, if that is so, then the CM, though supposedly exhibiting exemplary moral agency, would appear to be devoid of moral responsibility—a very odd result.

The short way with this objection would be simply to reject that libertarian contra-causal understanding of alternative possibility as irrelevant to the question of moral responsibility. This has been the traditional compatibilist maneuver. Unfortunately, no completely satisfactory competing account of alternative possibility has ever been given. Although compatibilists still make gestures in the direction of providing such an account, the matter remains very controversial. My strategy here is to try to meet the objection head on.

The objection rests on an assumption, challenged by Harry Frankfurt, 9 called the Principle of Alternative Possibilities: A person is morally responsible for what he has done only if he could have done otherwise. Frankfurt presented his own (controversial) counterexample to this principle. My method of reply to the objection is to adopt, for the moment, the libertarianism sense of 'can', and to take the present example of CM's cooperation as a very un-Frankfurt-like counterexample to the same principle. In this sense of 'can', the CM actor cannot do otherwise in the circumstances. I do not say that it is the primary sense of 'can' or 'could' which is relevant. I agree with Dennett that there should be an epistemic sense of 'possible' in which a person can treat various kinds of action as possible for her while she is deliberating (148). But that is compatible with there being another sense or use of 'possible' in which it may not be possible psychologically to do a certain thing. Sometimes an actor knows this. In his book, Peter van Inwagen quotes Mark Twain, who said: "I am morally superior to George Washington. He couldn't tell

⁹ "Alternate Possibilities and Moral Responsibility," this JOURNAL, LXVI, 23 (Dec. 4, 1969): 829–839.

Digitized by Arya Samaj Foundation Chennai and eGangotri

a lie. I can, and I don't" (63). Sure. But there are some things, even some lies, that a person knows he cannot go through with. This is the relevant sense.

To have a convincing counterexample to the principle, however, I need to make a convincing case that the CM in question is morally responsible, even though she cannot do otherwise. What is the argument? There are two stages. The first will not move the libertarian in the least, but it is essential background. This stage stipulates that, in our example, there is no coercion and no irrationality or limitation of freedom in the nature and formation of the agent's beliefs and desires. (The significance of this stipulation is the subject of the fourth part of this essay.) The second stage proceeds by reductio. It asks: What needs to be added to turn this paragon of instrumental rationality into a responsible agent? The correct answer, I believe, is: Nothing. But the libertarian will say: What is needed is the possibility, in the very strong sense indicated previously, of acting contrary to the CM disposition. But, I have argued, one cannot, in this strong sense, act contrary to the CM disposition. Hence, the libertarian free agent will be a SM maximizer and will defect. Since the present account of instrumental rationality implies that such action would not be rational (would not express a rational disposition to choose), there is the unacceptable consequence that we make an agent morally responsible by diminishing his rationality.10

(Note that the reductio does not rest on the assumption that a full expression of rationality is *necessary* for moral responsibility. There are many kinds of diminished rationality in which the agent remains morally responsible. Indeed, it is plausible to suppose that the conventional norms allowing cases of responsibility when there is diminished rationality can be morally justified by just the kind of appeal to instrumental rationality which implies the rationality of the CM disposition. This is, I believe, the strategy of argument in Dennett's discussion of responsibility in the last chapter of *Elbow Room*. The present defense of the CM's responsibility rests rather on the as-

¹⁰ If this is a counterexample to the principle of alternative possibilities, it is also a counterexample to an important version of van Inwagen's inference rule, "From $N(p \rightarrow q)$ and N(p), infer N(q)," where the N-operator applied to a proposition p results in the proposition that p is true and no one is, or ever has been, even partly rule is used to argue that determinism is incompatible with responsibility.) The relevance of the present example may be seen by letting P be a proposition describand by letting Q be a proposition stating that a person who has the CM disposition cooperates in the circumstances described earlier. If determinism is true (so that P reasons already given, N(Q) will be false.

sumption that undiminished rationality in a coercion-free context is *sufficient* for responsibility.)

III.

There are two worries of a general nature which have not been addressed. One is that the two-person, two-action, no repetition PD is too special and artificial to warrant substantial conclusions about free will. I am not sure that I have stated this objection in the best way. Clearly the compatibilist position is about logical possibility. It ought to be fair, therefore, to focus on a merely possible case. Moreover, the simplicity of the case should be a virtue, since it makes plain the logic of the argument. Still, one would think that the same argument ought to apply to more robust examples, and, if it did not, one would want to know why. The other concern is that the argument of the previous parts depends on the adequacy of Gauthier's conception of rationality. It is worth exploring the possibility that the reduction thesis might sustain a similar argument that is independent of Gauthier's specific understanding of instrumental rationality.

A way to make the PD itself more realistic is to allow for repeated interactions between the same persons. When we are in a position to cooperate with someone for mutual advantage but also are tempted to defect for an even greater advantage, one thing besides conscience which inhibits our defection is the prospect of being in the same kind of situation with this person again, plus the worry that others will know about the defection. So let us imagine agents who have a very high probability of meeting again in similar circumstances. Suppose the average number of replays between the same persons works out to be about 200. An immediate consequence is that the number of possible dispositions to choose jumps from a small number to two raised to the 200th power! Robust enough? Suppose also that there is a social context: some sixty persons, each of whom will interact with every other person in a series of PDs. Finally, let us imbue these sixty persons with various dispositions to choose which exhibit as much human intelligence as we can practically muster. We invite from around the world experts at strategic thinking to select dispositions from the trillions that are possible. We allow the experts randomizing devices, speedy computers to make calculations, a chance to experiment with different dispositions before settling on a choice of dispositions, and a chance to copy and improve on the success of others. As you may know, the experiment that I am describing is the one which has already been done by Robert Axelrod. 11 I shall cite here just a few features of his computer

¹¹ The Evolution of Cooperation (New York: Basic Books, 1984), chs. 1-3.

PD tournaments to indicate why I think my defense of compatibilism is sustained in this more realistic context.

The entries in one of Axelrod's round robin iterated PD tournaments came from artificial intelligence researchers, mathematicians. economists, and so on, from six countries. Sixty-two different dispositions to choose or decision rules (as Axelrod calls them) were used. including a purely random decision rule. Some rules were very sophisticated. For example, one of them produced Bayesian models for the other players, updating in Markov fashion its estimated conditional probabilities for the moves the other would make given its own, and then calculating the expected utilities of its options time by time on the basis of these conditional probabilities. As you probably know, the star of the show was an extremely simple rule called Tit for Tat, entered by the psychologist Anatol Rapaport. This rule says: Cooperate the first time and afterwards just copy the other player's previous move. Despite its utter simplicity and rigidity, Tit for Tat repeatedly achieved the highest utility score in this and other tournaments. In the sixty-two entry tournament, new contestants were made aware of Tit for Tat's previous success and given an analysis of the reasons for its success. Although they devised subtle new rules specifically designed to better Tit for Tat, none were able to.

This may seem surprising. Tit for Tat, by its very structure, cannot do better than its partner in any given series of moves. (It only defects after it is defected against, and immediately returns cooperation for cooperation.) In this respect, it is an analogue of Gauthier's CM rule.12 It is a maximizing rule, but one which maximizes in a constrained way in each PD interaction, only maximizing there to the extent that its partner maximizes. But it is also like CM in the respect that it is engineered to discourage defection against it; a consistent defector only gets one exploitation pay-off playing against Tit for Tat. Moreover, it is like CM in that it is easy to see ahead of time (as in Gauthier's second-order decision problem) that the best way to treat Tit for Tat is to cooperate with it. Axelrod's statistical analysis of millions of tournament interactions confirm that these features of Tit for Tat explain its success. When other rules in the tournament shared some of these features, there was a correlation between utility ranking of the rules and the extent to which they shared the CM-like

¹² Gauthier disagrees (p. 169, note 19), because "constrained maximizers may co-operate even if neither expects her choice to affect future situations," as in iterated PDs. But Gauthier allows their choice of dispositions to affect their future interaction, just as in Axelrod's tournament the agents' choice of decision rules affects their future interaction.

There can be no doubt, however, that the Tit for Tat "agent" is a paradigm of mechanism and utterly devoid of libertarian free will. To parallel the argument before, we need to consider what the tournament results might have been had the decision rules been replaced by agents with libertarian free will. Such agents would, in effect, make up their decision rules as they go along, deciding whether or not to cooperate move by move, unbound by any predetermined response pattern. Would they improve on Tit for Tat? It is impossible to give a definitive answer. There is a simple mathematical proof 13 to show that, if the probability of meeting the same individual again is high enough, there is no strategy that is best over all interaction environments; which pattern maximizes depends on the other patterns with which it is interacting. The most we can conclude is that there is no guarantee that metaphysically free agents would evince superior instrumental rationality, and that there is much indirect evidence that very smart human beings, at least, would initially do a lot worse.

But let us take this thought experiment a step further, giving the libertarian free agents the benefit of the doubt at every turn. It is at least possible that initially the tournament would proceed not very differently from Axelrod's experiments. Libertarian free actors might come up with Axelrod's proof that there is no absolutely best decision rule. They could then begin to experiment using decision rules that seem initially to be worth a try. After a while, perhaps a long while, when they have done Axelrod's statistical analysis, they could conclude that certain decision rules are better than others and begin switching to them. In fact, Axelrod simulated this experiment by running an ecological program in which the worst performing rules got eliminated and were replaced by the better ones. (The process can be given a Darwinian or a Lamarkian interpretation.) The result was that the more exploitative decision rules had a heyday (their success depending on having lots of easily exploitable strategies around to interact with) but eventually they too were driven to extinction and the population went to fixation at Tit for Tat. This seems a reasonable expectation for what might happen among smart (and not too stubborn) libertarian free agents as they switched from less successful to more successful patterns of decision. But would they stay at that point? Yes, they would, given the probability for repeated interaction and the PD pay-offs set in the tournament, for Axelrod has proved that, in this case, Tit for Tat is "collectively

¹⁸ Axelrod, op. cit., p. 15; see also his "The Emergence of Cooperation Among Egoists," The American Political Science Review, LXXV (1981): 309.

stable," i.e., a population of interacting Tit for Tats cannot be invaded by a new decision rule (59). Any libertarian free agent switching to an alternative pattern would start losing utility. This stage may be seen as an evolutionary solution to Gauthier's compliance problem.

The upshot is that, under charitable assumptions, we can expect that a community of smart libertarian free agents would end up doing as well as, but no better than, their rule-bound, metaphysically unfree, counterparts. Libertarian free will turns out to be, at best, perfectly idle. From the standpoint of instrumental rationality, the only variety of free choice worth wanting in the robust circumstances described is the freedom exemplified in the actions of Tit for Tat. The reduction thesis implies that the same can be said from the standpoint of morality. In particular, since libertarian free agency is inessential to being instrumentally rational, the reduction thesis via (RT2) supports the conclusion that libertarian free agency is inessential to being a moral agent.

IV

In this last part, I shall raise three difficulties for the reduction of morals to instrumental rationality. Although I believe that these problems are important and interesting in their own right, this is not the place to develop them in detail. As rational decision theory continues to evolve, it may in time be able to resolve these problems. ¹⁴ My interest in raising them now is twofold. I want to make the point (a) that their existence does not undermine the argument developed thus far against libertarianism, but at the same time I want to concede (b) that they reveal the possibility of serious limitations on freedom and that this possibility is masked by presuppositions on both sides of the compatibilism debate.

The first difficulty resides in the presupposition of rational choice theory that there is a reasonably well-defined set of outcomes in problems of strategic choice. This is a natural presupposition, for problems of strategic choice arise only when actions jointly make a difference to the agents. But how is the set of outcomes to be determined? It is to be determined, in part, by the sets of alternative actions that are possible, one set per agent. And how is one of those sets to be determined? Here rational choice theory does not have much to say.

¹⁴ The complex, interdisciplinary character of the pressures on rational choice theory to take account of the kinds of difficulty broached in this part is amply the difficulty for utilitarianism raised at the beginning of ch. 3 (which Elster calls reduction thesis.

On the assumption that the earlier defense of compatibilism is sound, it might be thought that compatibilist analyses of possibility are the appropriate candidates for extending rational choice theory in this direction. But this would be to misperceive the problem. Compatibilists like libertarians have generally focused their energies on understanding the ways in which a morally responsible agent can control which alternative is realized among some given set of conceivable options. The problem has been to determine in what sense the agent can make the future go one way rather than another. The background set of merely conceivable options is, as in rational choice theory, just given. The libertarian and compatibilist speak of choosing, for example, between taking money out of the poor box or refraining15 and of murdering Smith or refraining,16 as if it did not matter for their purposes whether there are more relevant ways to partition the conceivable options. But, even if we can all agree about how an agent can make one possibility actual instead of another, there is still the puzzle about how the set of conceivable options is identified in the first place. Moreover, figuring out the "relevant" alternatives, especially in problems of strategic choice, can be itself a matter for moral deliberation. If so, there is the danger that the program of reducing moral justification to instrumentally rational choice will degenerate into circularity.

The circularity problem is acute if the very institutions which are up for moral assessment and whose justification depends on people's hypothetical rational choices play a large role in shaping their conception of the feasible alternatives. Feminists have been particularly sensitive to this dimension of moral agency, and have argued that there tend to be fundamental and systematic differences in the ways men and women think about decision problems, especially with respect to the partition of possible choices.¹⁷ Such differences are completely masked by the rational choice model. Each person making a choice can think of a logically exclusive and exhaustive list of alternatives, and these sets of alternatives can be combined in strategic choice situations to form a more fine-grained partition. Rational choice theory is amenable, as it were, to anything, but at the same time it is of no moral help. If one does not think of a certain way of

¹⁵ van Inwagen, op. cit., p. 127.

¹⁶ Frankfurt, op. cit. Yrankfurt, op. cit.

17 See, for example, Carol Gilligan, In a Different Voice (Cambridge, MA: Harvard, 1982), and "Reply by Carol Gilligan," Signs, XI (1986): 324–333; Nel Noddings, Caring (Berkeley: University of California Press, 1984); Annette Baier, "Trust and Antitrust," Ethics, XCVI (1986): 231–260; and Kathryn Morgan, "Women and March March 1986. "Women and Moral Madness," Canadian Journal of Philosophy, Supplementary Volume XIII (1987): 201-226.

behaving and represent it in the choice set, it cannot, on this model. enter into one's rational deliberations. Yet thinking of a morally relevant alternative may be ninety-nine percent of coping with the decision problem. Ordinary agents will form a manageable partition of alternatives according to their sense of relevance, and their sense of relevance (in some cases, their conception of what is in their control) will reflect the influence of norms that are to be justified through rational choice.

The difficulty of identifying a morally relevant partition of alternatives is closely connected with a second problem for the reduction thesis. Just as rational choice theory takes as settled some set of alternatives, it takes as settled the agents' individual preferences over the set of possible outcomes. As is often stressed in textbook expositions of the theory, the theory is morally neutral with respect to the content of preferences, as long as they are consistent. At the same time, it is a commonplace in the social sciences that people's preferences are a product, in part, of a socialization process that includes the learning of moral values. It appears, therefore, that there is a second deep source of circularity in the reduction program.

The problem shows up in a dramatic way in Gauthier's bargaining problem. Even if this problem can be solved within its own terms so that all parties are satisfied that the bargain they have struck is thoroughly fair (or as good as their instrumental rationality will allow), there is nothing in the bargaining mechanism which permits the content of consistent preferences to be questioned. If preferences are systematically biased from the start in favor of the moral norms of the day, this bias will be tempered, if at all, only as an accidental by-product of the bargaining process. If, to give an obviously oversimplified example, men's apparent preferences by and large reflect a conception of themselves as superior in status to women, and if this conception tends to be shared by many women and reflected in their preferences as well, the bargaining solution cannot, by its own logic, reverse this conception. Indeed, the solution would imply that the social institutions that reinforce this prejudice are morally justified.

To be fair, the authors interested in reduction do pay some attention to the content of preferences. Gauthier stresses that the relevant preferences must be "considered," meaning that they must be fairly stable under relevant experience and reflection (22-40). Richard Brandt formulates and defends a conception of rational desire, which is roughly desire that is stable under relevant experience and reflection and which, in Brandt's theory of moral justification, plays a role similar to Gauthier's considered preferences. It might be thought, then, that truly considered preferences would be cleansed CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

of any distorting moral bias that would make them unsuited for a foundational role in moral justification. I believe that this thought does not accord with our evolved nature as social animals. The "relevant experience" and the social availability of information for reflection may, for a long time, serve only to reinforce the character of present preferences. This kind of point about the pervasive nature of social prejudice is stressed throughout the feminist literature, but of course not only there. It would appear that the integrity of individual preferences, however considered, cannot be assessed independently of assessing the social institutions through which they are created, sustained, and modified. Yet it is the social institutions themselves whose moral assessment is to be based ultimately on optimal agreement among individual preferences.

The problem of potential circularity which besets the reduction thesis with respect to outcomes and preferences is simultaneously a problem for free choice, but one which is masked by the focus, of libertarians and compatibilists alike, on the sense in which an agent can choose between genuine alternatives. Let the agent have the freedom (in whatever sense you like) to control which alternative in a given set will be realized. If the agent's conception of his alternatives or his preferences are under the control of others (through, say, a tiny Martian implant in his brain¹⁸ or by more familiar means), then his freedom of choice is drastically limited in ways that seem obviously undesirable. We expect a moral agent, in a sense robust enough to accord with commonsense, to be responsible to some significant degree for his conception of alternatives and for his considered preferences. But understanding his moral agency in these respects encompasses freedom that is not reducible to his being able to do otherwise or to his instrumental rationality, as that is currently conceived. At the very least, such an understanding would have to be built on a psychology of choice which is sufficiently naturalized to explain the "givens," or, as Herbert Simon¹⁹ calls them, the "boundary conditions," of present day rational choice theory.

Finally, I turn to a source of criticism more sweeping than the problems discussed so far. This is contained in Annette Baier's²⁰ recent indictment of the social contract tradition in ethics. Her dis-

¹⁸ Here I adapt for my purpose an example from van Inwagen (109); the implant interfers with the agent's free will, but in *this* case without determining his choice.

¹⁹ See "Human Nature in Politics: The Dialogue of Psychology with Political Science," *The American Political Science Review*, LXXIX (1985): 293–304.

Trust and Antitrust," op. cit.; see also her "What Do Women Want in a Moral Theory?" Nous, XIX (1985): 53–64. Similar criticisms of contractarian moral philosophy have been made by other authors, for example, Susan Sherwin in "A Feminist Approach to Ethics," Dalhousie Review, LXIV (Winter 1984/5): 704–713; and Virginia Held, "Non-Contractual Society: A Feminist View," Canadian Journal of Philosophy, Supplementary Volume XIII (1987): 111–137.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

cussion involves many interlocking themes, including a theory about the moral possibility of trust; I will extract one theme (from "Trust and Antitrust") to illustrate the relevance of her criticism. Although I present it separated from the web of Baier's supporting arguments, its force should be amply clear.

The criticism that I have in mind begins with her observation that the contract model of moral relationships—here we can add: especially the contemporary model that incorporates the reduction thesis coupled with current rational choice theory—depends on there being minimal trust between persons of comparable power. In cases of extreme disparity of power, for example, a healthy adult in relation to young children, a handicapped relative, the ill, the dying, or a dependent parent, the contract model has no direct application. In Gauthier's terms, when one party has no power leverage, there is no bargaining problem; and hence no problem of compliance with a rational bargain. Much of human experience, however, notably the moral experience of women, is bound up in ties of trust between persons who are significantly unequal in power. Such relationships, involving, for example, family and community, are typically unchosen and, in morally essential respects, cannot be contracted. Yet they constitute a primary sphere of moral assessment, choice, and responsibility. In brief, the reduction thesis, as it is spelled out in the contractarian model of moral relationships, fails to capture the full range (perhaps the core) of human moral experience.

This criticism may appear to be entirely external, applying an independent moral standard to judge the adequacy of the contractarian model, in contrast to the previous objections which raised an internal problem of circularity. But the present objection is not unconnected with circularity. Social institutions can be largely responsible for certain inequalities in power between persons. Institutions can also exacerbate or at least fail to compensate for inequalities that arise independently of them. To assess the moral legitimacy of an institution on the basis of a bargain that is already biased in its favor because of institutionalized inequalities between the persons affected is clearly question begging. It is not clear how question begging can be avoided within the limits set by the reduction thesis. For any instrumentally rational bargain, even one concerned with the issue of what the power inequalities should be, will itself presuppose some set of possible outcomes that will manifest differences in power. It is only by reference to some such outcome space that any instrumentally rational bargain can be struck.21

²¹ Gauthier anticipates some of this criticism, for example, in the final chapter of sponses in my critical notice (op. cit.).

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

In this part of the paper I have suggested a broad conception of free will. My will can be influenced through someone else's control of the context of choice—alternatives, preferences among outcomes, the set of possible outcomes—without interfering with the exercise of my will within that context. My suggestion is that a variety of free will worth wanting is one that includes some power to influence the variables that govern the context of one's choices. How much contextual freedom, as we might call it, is worth wanting or possible are questions I cannot pursue here. My concern in this part has been to point out how the issue of contextual freedom is masked within the framework of both the reduction thesis and the traditional debate over compatibilism.

If my arguments hold here, have I not undermined my defense of compatibilism? It should be clear that I have not. The complaints against the reduction thesis have concerned matters of contextual freedom which are illegitimately suppressed when morality is reduced to instrumental rationality (in roughly the way it is currently understood). But the earlier arguments focused on the traditional issue of free choice within a given context of choice, that is, where the alternatives, outcomes, and order of preferences are taken as given. Grant, for the sake of argument, that the reductionist theory of moral justification needs to be supplemented before it can satisfactorily explain how choices can be morally justified. The original defense of compatibilism should still hold as long as there remain contexts with the PD structure.

Might the PD context not be ruled out altogether in a fully adequate account of morally justified choice? This expectation is unlikely. Not only is minimal trust between equals not that rare, but more than minimal trust would be morally unjustified on many occasions. ²² Often enough in these cases the equilibrium point diverges from the optimum. Moreover, we need not be restricted to the single interaction PD. Life-size examples of iterated PDs, well documented in the literature, serve my argument just as well. In such circumstances, incompatibilist free will adds nothing to any choice worth wanting.

RICHMOND CAMPBELL

Dalhousie University

²² A criterion for identifying these occasions is given in Baier's "Trust and Anti-trust."

BOOK REVIEWS

The Philosophical Disenfranchisement of Art. ARTHUR C. DANTO. New York: Columbia University Press, 1986. xvi, 216 p. \$25.00; and The State of the Art. ARTHUR C. DANTO. New York: Prentice Hall Press, 1987. 228 p. \$19.95.

"You use a glass mirror to see your face," George Bernard Shaw once wrote; "you use works of art to see your soul." This sentiment, which Arthur Danto is fond of attributing to Hamlet, animates the body of his writings on art and aesthetics. These constitute some of the best art criticism as well as the most suggestive and exciting, if not yet complete, project in the philosophy of art in recent years. They are writings worth reading.¹

Andy Warhol's *Brillo Box* is visually indistinguishable from its numberless commonplace prototypes and yet only it, and none of them, is a work of art. A work of art, Danto has argued, is therefore never identical with its "material counterpart." The relationship involved here is extraordinarily complex, and Danto has been occupied with it since the publication of "The Artworld" in 1964. He has not yet given it, in my opinion, a definitive account, but he has made a convincing case to the effect that "the differences are as deep as those between bodily movements and actions. . . . The relationship between the work and its material substrate is as intricate as that between mind and body" (TC, pp. 100, 104). "I find myself," he now writes, "wanting to think of *Brillo Box* as possessing a soul, admittedly imperceptible since its counterparts, lacking souls, look exactly like it" (PDA, p. 179).

The visual identity of some works of art with objects that are not themselves artworks is the founding intuition of Danto's philosophy of art, and underlies some of his most original and controversial views. It has, for example, remarkably novel implications for the traditional problem of the definition of art, which has been both the staple and the bane of most of the literature in aesthetics.

Since some artworks cannot be visually distinguished from mere things, Danto argues that the class of artworks will never be isolated by an appeal to any of its sensory features: nothing we can see will

¹ In parenthetical page references below, I shall abbreviate the title of *The Transfiguration of the Commonplace* (Cambridge: Harvard, 1981) as TC, that of *The Philosophical Disenfranchisement of Art* as PDA, that of *The State of the Art* as SA.

² This JOURNAL, LXI, 19 (October 15, 1964): 571–574.

ever explain the difference between artworks and ordinary objects. On the one hand, actual cases that show this are provided by *Brillo Box*, by Marcel Duchamp's readymades, and by Roy Lichtenstein's giant cartoon frames. On the other, for any such features we may imagine, Danto, the grand master of the no-difference that makes a difference, is ready to describe an object possessing them but failing to be a work of art.

This is not to say that Danto believes that aesthetic features (things like grace, delicacy, daintiness, handsomeness, comeliness, elegance, garishness)³ are not important to artworks. But he does believe—and this is one of his view's main strengths—that aesthetic features, which are supposed to be sensory or quasi-sensory, cannot support a definition of art. On the contrary, only after an object has been taken as an artwork on independent grounds can we tell which of its many sensory (and nonsensory) features are aesthetically relevant to it: "Aesthetics does not pertain to the essence of art" (PDA, p. 31).⁴ A brilliant application of some aspects of this idea occurs in Danto's devastating review of the show on Primitivism and 20th Century Art at the Museum of Modern Art (SA, pp. 23–27).

What, then, pertains to the essence of art? Danto's current answer, not always clear in its details but deeply promising, is that works of art are constituted by interpretation. Duchamp did not merely classify his urinal as an artwork, he interpreted it as such. Interpretations, accordingly, are "functions which impose artworks onto material objects, in the sense of determining which properties and parts of the latter are to be taken as part of the work and within the work significant in a way they characteristically are not outside the work" (PDA, p. 42). Such constitutive interpretation, Danto believes, is indispensable, and a critic's interpretation is correct when it "coincides most closely with the artist's own interpretation" (PDA, p. 44).

More needs to be said about the relationship between artworks and their material counterparts, about the general features that distinguish interpretation from other attempts at explanation, and about the sense in which interpretations are "imbedded" in art-

³ This is one of the lists provided by Frank Sibley in "Aesthetic Concepts," in George Dickie and R. J. Sclafani, eds., *Aesthetics: A Critical Anthology* (New York: St. Martin's, 1977), p. 817.

⁴ Danto's approach can be usefully compared with the view of Nelson Goodman [presented, among other places, in "When Is Art?", in his Ways of Worldmaking (Indianapolis: Hackett, 1978), pp. 57–70] that symbolic properties, which cannot be construed as sensory, are relevant to the artistic status of things. Goodman believes that the presence of such properties can explain the fact that some objects function as artworks.

works. But I prefer to move on to another application of Danto's fundamental idea.

Let us ask again why *Brillo Box* is an artwork while its look-alikes are not. The answer is given by an interpretation according to which *Brillo Box* (and none of its look-alikes) poses for its viewers the question why *Brillo Box* is an artwork while its look-alikes are not. This marvelous question presupposes that *Brillo Box* is an artwork and therefore, simply in virtue of being asked, establishes it as one. But the question is also none other than the philosophical question of the definition of art. *Brillo Box* declares itself to be an artwork simply by asking how this can be possible!

For Danto, this was an unprecedented event. It constituted, in his eyes, nothing less than the end of art, an issue he pursues with imagination and eloquence both in his criticism and in his philosophical essays. Very briefly, his view is this. For a long time, art took itself to be engaged in a progressive enterprise aiming at the perfectly faithful representation of reality. But the cinema dashed these hopes. "Moving pictures" duplicated and did not merely represent motion and thus showed that perfect fidelity to the world was impossible for painting and sculpture.

Faced with this crisis, modernist art then turned to an investigation of what it was, after all, that it was doing. Each of the many movements characteristic of the first half of our century offered its own answer to that question until *Brillo Box* presented just that question as its subject. With this, "art was taken up into philosophy, since the question it raised and the form in which it was raised was as far as art could go in that direction—the answer had to come from philosophy. And in turning into philosophy, one might say that art had come to a certain end" (SA, p. 209).

Paintings, of course, will continue to be made; what we cannot expect any longer are breakthroughs. Danto longs for the seriousness of modernism (does modernism include, one wonders, Warhol or Robert Rauschenberg?), when "each definition was accompanied by a severe condemnation of everything else as not art. There was an almost religious fervor . . . like the strife of warring sects" (SA, p. 217). By contrast, the carefree pluralism of the 1970s showed that "when one direction is as good as another direction, there is no longer any concept of direction to apply" (PDA, p. 115). As to the Neo-Expressionism that has swept the 1980s, it is only a recycling of old ideas, now bigger and worse, fuelled by the art market (SA, pp. 43–47, 77–80). In any case, it is not new. "The time for next things is past. . . . It was like coming to the end of the world with no more continents to discover. One must now begin to make habitable the

only continents that there are. One must learn to live within the limits of the world" (SA, p. 217).

It is always difficult to know how to argue with announcements, either in celebration or in lamentation, of the end of an era. If we implicitly identify the world with the earth, then Danto's striking simile seems perfectly appropriate. If, however, we consider what a small part of the universe the earth actually is, the continents' completed exploration is no longer the end of the world. Danto's narrative, moreover, is often selective: politically directed modernist movements find no place in his account. Nevertheless, it is a complex and unsettling narrative, and it inspires much of his superb criticism. Since it has already provoked a series of responses,⁵ I want to leave it behind and turn to one of the activities in which art, according to Danto, has engaged in its "post-historical" era. This is an activity he calls "disturbation," and leads quickly to some important issues.

Disturbatory art is not art which merely represents disturbing realities. Although Danto does not explicitly discuss this, we can see disturbatory art as a response to the question *Brillo Box* raised concerning the distance between art and reality. Such art aims to disturb by eradicating that distance, by literally incorporating reality—untransfigured, as it were—into artworks. It thus aims to frustrate and unsettle its audience's aesthetic, distanced, and contemplative expectations: "Reality must in some way . . . be an actual component of disturbatory art, and usually reality of a kind itself disturbing. . . And these as components in the art, not simply collateral with its production or appreciation" (PDA, p. 121). Happenings, for example, were disturbatory and so are Chris Burden's viciously self-endangering projects.

Disturbatory art, for Danto, is atavistic, aiming to bring reality back into art as, we are to suppose, once was the case: "Once we perceive statues as merely designating what they resemble . . rather than containing the reality through containing the form, a certain power is lost to art" (PDA, p. 128). Danto sees this as an effort to restore art to an earlier, "magical" stage when the object represented was thought literally to be present to its audience. But he confesses: "I do not enjoy disturbatory art, perhaps because I am always outside it and see it as pathetic and futile" (PDA, pp. 132/3).

But why should we expect to enjoy an art form primarily intended to question the very idea that art is to be enjoyed? The reason, I suspect, is that Danto insists on identifying art in general with the fine arts, to which enjoyment (wrongly, in my opinion) has always

⁵ See Berel Lang, ed., The Death of Art (New York: Haven, 1984).

seemed central. More important, I think, disturbatory art is necessarily bound to be "pathetic and futile." Such art depends essentially on the forms and conventions of the fine arts, whose "fineness" depends precisely on the distance they have managed to insert between representation and reality, and this distance can no longer be bridged. In any case, disturbatory art is always disarmed "by cooptation, incorporating it instantaneously into the cool institutions of the artworld" (PDA, p. 119) not in order to counteract its disturbing aspects but simply because this is where it has always belonged.

The disturbatory artist tries to restore "to art some of the magic purified out when art became art" (PDA, p. 131), to restore to fine art some of the immediacy it was felt to have had when it was still a living, popular genre. I want to claim that this is not a reasonable goal: fine art cannot be, at least in this sense, defined. This is a point Danto seems to accept, without remarking on the parallel, in his reaction to a mound of tires exhibited outside the Whitney Museum: "We are all prepared to see tires elevated to art, but there is a barrier against seeing paintings demoted to the status commonly occupied by tires" (SA, p. 10).

And yet "the magic purified out when art became art" is all around us, though it is exactly for that reason almost totally invisible. The distinction between representation and reality is successfully blurred by television—literally an art which has not become art—and which truly disturbs its audience. As a medium, television is still transparent in that it makes us believe that what we see in it is precisely what we see through it. Its magic may be neither admirable nor even respectable. But it is structurally identical to the magic Plato saw in Attic drama, which also, of course, was not art. Plato's disenfranchisement of poetry, therefore, may not have been a disenfranchisement of art but a disenfranchisement of popular culture.

This disenfranchisement, I believe, is constantly being duplicated today. We find it not only in the various attacks on television on the part of cultural critics but also in the total and absolute neglect of it on the part of our philosophy of art. The arts Danto wants to reenfranchise (PDA, p. 169) are just the arts which can no longer do much harm. Every literary work, he writes, is "about the 'I' that reads the text... in such a way that each work becomes a metaphor for each reader" (PDA, p. 155). The key word here is 'metaphor': we do not literally emulate our literary heroes, which was just what Plato was afraid would happen with the poetry of his time and just what so many of us are afraid of in connection with television. Looked at from the point of view of what it excludes, Danto's philosophy of art can be seen as part of the Platonic project.

But what a glorious part it is. Danto's critical pieces—witty and urbane essays—are uniformly a joy to read. His philosophical writings, which concern many issues which I have not even managed to mention here, bristle with ideas and suggestions and are gradually falling in place as elements in a grand and sweeping system. It would be appropriate for Danto, like posthistorical painters, to turn to a careful mapping of the interior of the continents he has explored; "God is in the details," Danto quotes Flaubert saying: we need more of those. But I also wish he would try to explore the new continents provided by the popular arts. These arts, too, are part of the domain of the philosophy of art and, for better or worse, they too provide us with mirrors in which to catch not only the conscience of our kings but also perhaps our own.

ALEXANDER NEHAMAS

University of Pennsylvania

The Truth in Painting. JACQUES DERRIDA. Geoff Bennington and Ian McLeod, trans. Chicago: University of Chicago Press, 1987. xiv, 386 p. \$27.50 cloth, \$19.95 paper.

Derrida's book has five parts: introductory remarks about truth in painting; catalogues of two eminently forgetable visual artists, Adami and Gerard Titus-Carmel; an essay, "Parergon," on Kant's third critique; and a discussion of Meyer Schapiro's critical note on an example in Martin Heidegger's The Origin of the Work of Art. The central topic is the impossibility of definitely establishing what is and what is not inside the frame. Although Derrida says in what I call the introduction that he "cannot report what is going on," in the next sentence he seemingly does so; such a summary would, he says, introduce into his narrative itself "the very indecision which I was trying to reduce" (2). A summary mentions the main points, while omitting those which are less significant. For example, normally a summary of the third critique presents Kant's theory of aesthetic judgment, leaving aside his discussion of wild flowers; a summary of The Origin of the Work of Art discusses Heidegger on equipment, but not his brief mention of a Van Gogh painting of shoes. But Derrida's deconstructive account of Kant and Heidegger focuses, characteristically, on what traditional commentators would judge marginal elements in their work.

0022-362X/88/8504/0219\$00.50

© 1988 The Journal of Philosophy, Inc.

More exactly, Derrida questions this very distinction between what is marginal and what is central in a book. Kant quotes Monsieur de Saussure, in whose Journey in the Alps we can "read something that Kant did not quote . . ." (85), but which Derrida will discuss; Van Gogh's pair of shoes leads us to another pair not mentioned by Heidegger, those in Van Eyck's Arnolfini Marriage. Derrida often criticizes others for quoting out of context. Schapiro "cuts out of Heidegger's long essay, without further precautions, twenty-odd lines, snatching them brutally from their frame which Schapiro does not want to know about and then interpreting them with . . . tranquillity . . ." (285).

But, in summarizing Derrida's argument, what am I doing if not also cutting passage out of context, snatching his words brutally from their frame? Indeed, since *The Truth in Painting* quotes *Glas*, other works by Derrida, Kant, Hegel, and others, how could it be reviewed without taking words out of context? The translators' Preface says, "The 'compromise English' in which this translation is written . . . should stand in no need of excuse" (xiv) and everyone would agree, quoting that as—"The 'compromise English' in which this translation is written . . . should stand in . . . need of excuse" is to quote out of context. But, since what they are offering, arguably, is an excuse for their "compromise English," one needed, perhaps, since their translation contains many discussions of Derrida's French, this "unfair" quotation may well represent their intentions.

How plausible are Derrida's arguments? I focus on the essay I am best qualified to evaluate, the discussion of Van Gogh. Heidegger, like many speculative philosophers, simply picked an inapt example, confusing an image of the artist's shoes with those of a peasant woman; and Schapiro had reason enough to criticize that detail in an essay in Festschrift for a refugee from the Nazis. Derrida's account is more complex. Heidegger is "ridiculous and lamentable" (292); his error is to confuse what is in the picture with what is represented in it. But Schapiro's account is no better. He fails to understand the role of such an example in Heidegger's aesthetic (295, 297); mistakenly attributes "to Heidegger an intention, which nothing in his text shows to be there, of describing and referring only to painted shoes" (320); and wrongly asserts that, because Van Gogh was not a peasant, but a man of the city, he could not identify with a peasant (362). Derrida's interesting suggestion is that the larger context influences a writer's conclusion about the identification of represented objects; where Heidegger sees these shoes as equipment belonging to the earth, Schapiro finds them the property of a sophisticated citydweller. The dull claim, which Derrida seems also to make, is that the identity of the shoes is indeterminant. But, just as the claim that either statements have absolutely determinant meaning or no meaning makes interpretation impossible, so this similar point about depicted objects gives an unconvincing argument against a familiar procedure of art historians. This is but a variation on traditional skepticism. Instead of arguing, "we can never know whether there is an external world," Derrida says; "we can never know whether our statements (or pictures) have determinant reference." But this is to set an unrealistically high standard for art historical evidence.

Yet, to treat Derrida thus as offering debatable arguments fails to do justice to his style which cannot, I believe he says, be separated from his argument. His essay on Kant is broken up into a sequence of partially framed segments; his account of Schapiro is a "polylogue" composed for "n + 1 - female - voices" (256): and these features of his writing are as essential to his books as Wittgenstein's use of examples or Nietzsche's of aphorisms are to theirs. How then should awareness of his style inform philosophers' argumentation with Derrida? The existing literature offers two models.2 We may domesticate Derrida, treating him as a fellow philosopher who argues in an "odd" style. Alternatively—as literary critics suggest we can treat him as a stylist. If we accept the first model, then our goal—as in my discussion in the previous paragraph—is to extract Derrida's arguments from their context. Yet, as we might expect from a professor of philosophy whose constant theme is the critique of any dualisms—the distinction between what is inside and what remains outside the frame is merely one variation on his theme—neither of these models can be entirely adequate. So, if sometimes he analyzes the style of philosophical works, "treating the third Critique as a work of art or a beautiful object . . . as if the existence of the book were indifferent to me (which, as Kant explains, is a requirement of any aesthetic experience)" (43), even if, he adds, Kant did not intend this; elsewhere, as when he says, against Schapiro, that we cannot prove "that these shoes were those of Van Gogh" (365), he gives an argument. No analysis which fails to take account of both his style and arguments can do justice to his book.

¹ I have built on this argument in my "Manet and his Interpreters," Art History, VIII, 3 (September 1985): 320-335; and in my forthcoming Representations: The Tropics of Art Historical Discourse.

² As its title indicates, Henry Staten's Wittgenstein and Derrida (Lincoln: Nebraska UP, 1984) treats Derrida as a philosopher; as its title suggests, Stephen W. Melville's Philosophy Beside Itself: On Deconstruction and Modernism (Minneapolis: Minnesota UP, 1986) is more ambivalent about his place within philosophy.

Alexander Nehamas argues that Derrida holds two views, one true: "any statement can function in indefinitely many contexts"; one false: "no statement has a determinate meaning." How, he asks, can Derrida state these views without contradicting himself? Since this is as plausible a reconstruction as any of many of Derrida's arguments. note its difficulties. Is not the goal of Derrida's writing to show that any statement can function in any context? How, indeed, does he get from the discussion of frames to Kant's account of the tulip (81/2)?: why bring in "the phallus which doubles the corpse" (126) in discussing Kant on the colossal?; what motivates the transition from Heidegger's account of van Gogh's shoes to Magritte's shoes, fetishism and "castration as the truth of truth" (334)? The key, always, is his transitions. Derrida is always able to keep writing because he makes arbitrary transitions. Like the professor who, finding that she has arrived in class without her notes on Donald Davidson, improvises an account about W. V. O. Quine, Samuel Beckett, and her sailboat trip, Derrida makes no distinction between discussing the announced subject and discussing some other subject. Since Adami drew a self-portrait containing an image of Walter Benjamin, Derrida discusses "the incredible epos of a Jewish hero, of a revolutionary thinker" (181); Adami made a Study for a Drawing after Glas, so Derrida also mentions Hegel and Genet; since Schapiro wrote on the semiotics of frames, there is a narrative link between Kant's account of framing and Heidegger's discussion of Van Gogh: and so on forever. If, as in these examples, the meaning of a statement is determined by its context, then the meaning of any statement is indeterminant. Everything Derrida can talk about is linked with everything else, not because—as idealists think, "the world is one" but simply because there is no text which cannot be linked with another. Of course, every text contains transitions, but, by making his so obviously arbitrary—by concentrating on what traditional interpreters would think marginal elements in the texts of Kant and Heidegger-Derrida focuses attention on the role of transitions as such.

This book is very difficult to read, and—what is worse—extraordinarily boring. Some pages from the end, I note: "here I give up." Since there is, alas, little correlation between philosophical originality and literary skill, many important philosophers are very difficult to read. But Derrida is boring because his aim is to keep his readers

³ See his "How to Understand Jacques Derrida. Truth and Consequences," *The New Republic* (October 5, 1987): 31–36. My critique of this analysis is, ironically, indebted to arguments Nehamas has presented elsewhere.

waiting, ready to identify his transitions which mark his changes of topic. The "objective correlate" of this waiting is boredom. The unity of a text, Derrida's *style* shows, is always ultimately the product of conventions about an organic whole; and this perhaps constitutes an *argument* against Nehamas's reconstruction of Derrida's position. But, if I have correctly identified the internal connection between Derrida's style and his argument, then it seems that his argument is ultimately unexciting.

So much for the philosophical analysis. Were I a literary critic, I might imitate Derrida. The back cover shows him in a strange pose. Is he rising or kneeling down? He will end up, it has been said, a religious man. He is framed between two columns of unequal width, before a strangely unreal landscape. "In the interval between the mathematical sublime and the dynamic sublime," he writes, "a tree had been projected into the Milky Way" (146). Does not this photo illustrate this passage? This image comments on framing, columns, and pictorial reference, and its credit to "Donoso/Sygma" is—who can doubt that?—an anagram which more erudite reviewers will easily decipher. Unconcerned with the photographer's intentions, or Derrida's, but for the constraints established on reviewers by the editors of this journal, I could write an even longer text than *The Truth in Painting* on the relation between this photo and the book on which it appears.

DAVID CARRIER

Carnegie Mellon University

ERRATUM NOTICE

In Stephen Schiffer's review of Gareth Evans's *The Varieties of Reference* in our January issue, part of a sentence was inadvertently left out of a paragraph on page 40. The full paragraph is reprinted below. The editors regret the error.

I wish to call attention to the way in which thoughts involving a demonstrative idea, or mode of identification, of an object, such as a thought whose content is the proposition This was admired by many, is explained in terms of another thought about the same object which also involves a demonstrative idea of it, in this case a thought whose content is a proposition of the form This = the object at π now. This illustrates my earlier complaint about Evans's failure to provide a non-question-begging completion for the schema [*], but I want to use it now to raise a different problem for Evans's account of belief content in terms of modes of identification.

NOTES AND NEWS

The editors wish to remind our correspondents that notices of scheduled meetings and symposia are of value to our readers only if they appear early enough to make attendance possible. Copy for Notes and News must be planned and processed in this office in time for it to go to the printer at least six weeks before publication date, the 15th of each month. Accordingly, an announcement of a meeting to be held, say, in April, should be scheduled for our February issue (or sooner) and should be received in this office no later than December 1, that is, at least four months in advance. The editors appreciate notices that come to us ready for the printer, requiring little editorial work and no retyping, in Journal style (see these columns), with no imperatives, no prices, no telephone numbers, no headings, no unnecessary capitals; and reasonably terse, since space is limited.

The Austrian Wittgenstein Society will sponsor the Thirteenth Annual Wittgenstein Symposium on August 14–21 in Kirchberg/Wechsel. The theme of the symposium will be "Philosophy of the Natural Sciences." Sessions will include: Wittgenstein and his Attitude toward Science, Foundations and Philosophy of Physics and Chemistry, Foundations and Philosophy of Biology and Genetics, Cosmology, General Methodological and Interdisciplinary Problems and Perspectives, and History of the Natural Sciences. Scholars wishing to present a paper on any of these topics may request an abstract style sheet from the Society, A-2880 Kirchberg/Wechsel, Markt 2, Austria. Further information may be obtained by writing Philip Hugly, Phil. Dept., Univ. of Nebraska, Lincoln, NE 68588. Students interested in earning up to six graduate or undergraduate academic credits in conjunction with attendance at this symposium may write to R. L. Burr, Univ. of Southern Mississippi, MS 39406-5015.

The Husserl Archive at the University of Freiburg/Breisgau, West Germany, will sponsor a series of lectures on the occasion of the 50th Anniversary of Edmund Husserl's death. Bernhard Waldenfels (Bochum) will speak on May 5; Ernst Wolfgang Orth (Trier) on May 19; Bernhard Rang (Freiburg-Bonn) on June 9; and Klaus Held (Wuppertal) on June 23. Further information about these lectures and other events commemorating Husserl's death may be obtained by writing to the Husserl Archive, Univ. of Freiburg/Breisgau, D-7800, BRD.

The editors are pleased to announce that the International Hobbes Association will sponsor a new journal, *Hobbes Studies*. Editorial inquiries may be directed to Martin A. Bertman, Phil. Dept., Ben Gurion Univ., Box 653, Beer Sheva, Israel. Subscription orders may be sent to Van Gorcum Publishers, Box 43, 9400 AA, Assen, Netherlands.

NEW BOOKS: TRANSLATIONS AND NEW EDITIONS

ARISTOTLE: De Anima. HUGH LAWSON-TANCRED, trans. New York: Penguin, 1986. 254 p. Paper \$5.95.

BLOCH, ERNST: The Principle of Hope. NEVILLE PLAICE, STEPHEN PLAICE, and PAUL KNIGHT, trans. Cambridge, MA: MIT Press, 1986. 3 volumes: xxxiii, 447 p.; viii, 924 p.; viii, 1420 p. \$95.00 for 3-volume set. The Utopian Function of Art and Literature: Selected Essays. JACK ZIPES and FRANK MECKLENBURG, trans. Cambridge, MA: MIT Press 1987. xliii, 310 p. Cloth \$25.00.

BLUMENBERG, HANS: The Genesis of the Copernican World. ROBERT M. WALLACE, trans. Cambridge, MA: MIT Press, 1987. xlviii, 772 p. Cloth \$40.00.

BRANDT, WILLY. Arms and Hunger. ANTHEA BELL, trans. Cambridge, MA: MIT Press, 1987. vii, 208 p. Paper \$7.95.

CARNOIS, BERNARD: The Coherence of Kant's Doctrine of Freedom. DAVID BOOTH, trans. Chicago: University Press, 1987. xv, 174 p. Cloth \$23.95.

CASTORIADIS, CORNELIUS: The Imaginary Institution of Society. KATHLEEN BLAMEY, trans. Cambridge, MA: MIT Press, 1987. 418 p. Cloth \$35.00.

DERRIDA, JACQUES: The Postcard: From Socrates to Freud and Beyond. ALAN BASS, trans. Chicago: University Press, 1987. xxx, 521 p. Cloth \$46.00, paper \$18.95. The Truth in Painting. GEOFF BENNINGTON and IAN McLEOD, trans. Chicago: University Press, 1987. xiv, 386 p. Cloth \$49.95, paper \$19.95.

DESCARTES, RENÉ: Meditations on First Philosophy. JOHN COTTINGHAM, trans. New York: Cambridge, 1986. xvii, 120 p. Cloth \$19.95, paper \$6.95.

FREGE, GOTTLOB: Collected Papers on Mathematics, Logic, and Philosophy. BRIAN McINNESS, ed., MAX BLACK et al., trans. New York: Blackwell, 1984. 412 p. Cloth \$39.95.

HABERMAS, JÜRGEN: The Philosophical Discourse of Modernity: Twelve Lectures. FREDERICK LAWRENCE, trans. Cambridge, MA: MIT Press, 1987. xx, 430 p. Cloth \$27.50. The Theory of Communicative Action, Volume Two: Lifeworld and System: A Critique of Functionalist Reason. THOMAS A. McCARTHY, trans. Boston: Beacon Press, 1987. 457 p. Cloth \$37.50.

HEIDEGGER, MARTIN: Nietzsche: Volume III: The Will to Power as Knowledge and as Metaphysics. San Francisco: Harper & Row, 1987. xiii, 288 p. Cloth \$20.95.

KANT, IMMANUEL: Critique of Judgment. WERNER S. PLUHAR, trans. Indianapolis, IN: Hackett, 1987. cix, 576 p. Paper \$12.50.

LEVINAS, EMMANUEL: Collected Philosophical Papers. ALPHONSO LINGIS, trans. Boston: Nijhoff, 1987. xxxi, 186 p. Cloth \$56.50.

LORENZEN, PAUL: Constructive Philosophy. KARL RICHARD PAVLOVIC, trans. Amherst: Massachusetts UP, 1987. x, 291 p. Cloth \$30.00.

MARCUSE, HERBERT: Hegel's Ontology and the Theory of Historicity. SEYLA BENHABIB, trans. Cambridge, MA: MIT Press, 1987. xliii, 360 p. Cloth \$25.00.

MARX, WERNER: Is There a Measure on Earth? Foundations for a Nonmetaphysical Ethics. THOMAS J. NENON and REGINALD LILLY, trans. Chicago: University Press, 1987. xi, 172 p. Cloth \$22.95.

PELLEGRIN, PIERRE: Aristotle's Classification of Animals: Biology and the Conceptual Unity of the Aristotelian Corpus. ANTHONY PREUS, trans. Berkeley: California UP, 1987. xiv, 235 p. Cloth \$37.50.

PHILOPONUS: Against Aristotle on the Eternity of the World. CHRISTIAN WILDBERG, trans. Ithaca, NY: Cornell, 1987. 182 p. Cloth \$29.95.

PÖGGELER, OTTO: Martin Heidegger's Path of Thinking. DANIEL MAGUR-SHAK and SIGMUND BARBER, trans. Atlantic Highlands, NJ: Humanities. 1987. xix, 293 p. Cloth \$45.00.

REALE, GIOVANNI: A History of Ancient Philosophy: Volume I: From the Origins to Socrates. JOHN R. CATAN, ed. & trans. Albany, NY: SUNY Press. 1987. xxv, 425 p. Cloth \$44.50.

SCHÜRMANN, REINER: Heidegger on Being and Acting: From Principles to Anarchy. Bloomington: Indiana UP, 1987. viii, 406 p. Cloth \$37.50.

SIMMEL, GEORG: On Women, Sexuality, and Love. GUY OAKES, ed. & trans. New Haven, CT: Yale, 1987. viii, 194 p. Paper \$7.95.

STAMBAUGH, JOAN: The Problem of Time in Nietzsche. JOHN F. HUMPHREY, trans. Lewisburg, PA: Bucknell UP, 1987. 226 p. Cloth \$29.50.

STRAUSS, LEO: Philosophy and Law. FRED BAUMANN, trans. Philadelphia, PA: Jewish Publication Society, 1987. xvi, 138 p. Paper \$12.94.

A new volume in the Arizona Colloquium in Cognition-

Contents of Thought

Robert H. Grimm & Daniel D. Merrill, eds.

This new collection of essays offers a state-of-the-art overview of the role of the content of propositional attitudes in psychological theory.

> Cognitive Suicide/Lynne Rudder Baker Comments by Charles Chastain . . . Reply

The Explanatory Role of Content/Fred Dretske Comments by Robert Cummins...Reply

Cartesian Error and the Objectivity of Perception/Tyler Burge Comments by Robert Matthews . . . Reply

Social Content and Psychological Content/Brian Loar Comments by Akeel Bilgrami ... Reply Belief Attribution and Content/Robert Stalnaker

Comments by Hans Kamp 210 pages, May 1988, \$39.95. Contents of Thought represents the

Proceedings of the twenty-fifth annual Oberlin Colloquium in Philosophy.

Also in the Arizona Colloquium-

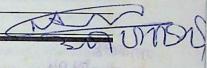
The Representation of Knowledge and Belief Myles Brand & Robert M. Harnish, eds.

368 pages, 1986, \$39.95.

The University of Arizona Press 1230 N. Park #102, Tucson, AZ 85719 CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 5, MAY 1988



THE REFUTATION OF INDETERMINACY*

WELL-CHOSEN title, unlike a name, informs us about its bearer. The title "Über Sinn und Bedeutung" was wellchosen, and so too was the title Word and Object. Gottlob Frege's essay claims that words have sense as well as reference. W. V. O. Quine's book denies precisely this. Its counterclaim is eloquently expressed by the conspicuous absence of a term translating "Sinn" in its title.

It is now widely thought that Word and Object, together with Quine's earlier writings in From a Logical Point of View,³ establishes his claim that the traditional intensionalist's notions of sense, synonymy, and analyticity cannot be made objective sense of, and, consequently, must be abandoned in serious studies of language. Accordingly, these works have been a watershed for twentieth-century Anglo-American philosophy, radically changing how philosophers think about language, logic, and nearly every other area of investigation. Quine's skepticism, especially as expressed in his indeterminacy thesis, have all but eliminated intensional approaches to language from the current philosophical scene.

On the metaphysical side, Quine's arguments were instrumental in resurrecting philosophical naturalism. Frege had all but single-

Gottlob Frege, "On Sense and Reference," in Translations from the Philosophical Writings of Gottlob Frege, Peter Geach and Max Black, eds. (Oxford: Blackwell, 1952), pp. 56-78.

² W. V. O. Quine (Cambridge: MIT, 1960). Hereafter WO. Cambridge: Harvard, 1953. Hereafter LPV.

^{*} The author wishes to thank Paul Horwich, Peter Lupu, David Rosenthal, Virginia Valian, and an anonymous reader for this JOURNAL for helpful suggestions. Earlier versions of the paper were presented at the Pacific Division Meeting of the American Philosophical Association, March 1986, in the symposium Word and Object: Twenty Five Years Later, and at philosophy colloquia of the University of Rochester and the City University of New York's Graduate Center.

handedly stemmed the tide of nineteenth-century naturalism in the philosophy of language, logic, and mathematics. 4 Carnap incorporated Frege's achievement into logical empiricism, giving that philosophy a strong non-naturalist orientation. Frege's sharp analytic-synthetic distinction, as explicated in Carnap's formal semantics, gave abstract objects and necessary truths sanctuary on the analytic side of the distinction. The distinction stood as the principal barrier to a return of an uncompromising naturalism in the spirit of J. S. Mill. This is why Quine, whose sympathies are clearly with empiricism, sets out to attack an empiricism in "Two Dogmas of Empiricism" (LPV. pp. 42-46). The arguments in that paper and his deployment of the indeterminacy thesis against the possibility of identity conditions for senses were widely seen as bringing this barrier crashing down and thereby opening the way for a neo-Millian naturalism of the sort sketched in "Carnap on Logical Truth." Hence, when the linguistic turn shifted the emphasis in philosophy to language, much subsequent philosophy came to be done within a naturalistic framework which might be described as Humean epistemology minus the category of relations of ideas. Matters of fact are all that matter.

Given the momentous changes Quine's arguments have brought about, it is desirable that they be subjected to careful and continuing scrutiny. In this paper, I examine the argument for indeterminacy from a new angle and find that it does not work. If I am right, there is a straightforward sense in which the indeterminacy thesis is refuted. Skepticism about translation, like skepticism about other things of which common sense assures us, incurs a burden of proof in challenging a common-sense view. If the skeptic provides reasons of sufficient strength to discharge the onus of proof, we are presented with an advance in knowledge whose surprising character marks it as a discovery of the most profound sort. Thus, we attach the importance we do to Quine's argument in large part because it threatens to upset our common-sense view that there is always a right and a wrong translation, even when the options differ the way "rabbit," "rabbit stage," and "undetached rabbit part" do. But, if the skeptic's reasons lack the strength to discharge the onus of proof and can establish only the logical possibility that common sense is wrong, we are presented with nothing more than an "absolute skepticism" which,

⁶This notion is adapted from "The Refutation of Idealism," in G. E. Moore, *Philosophical Studies* (Boston: Routledge & Kegan Paul, 1922), p. 30.

⁴ Hans Sluga, Gottlob Frege (Boston: Routledge & Kegan Paul, 1980), pp. 17-34. ⁵ In The Philosophy of Rudolf Carnap, Paul A. Schilpp, ed. (LaSalle, IL: Open Court, 1963), pp. 385-406; and also, of course, at the end of "Two Dogmas of Empiricism" in LPV.

applying to all forms of knowledge, proves too much. In this instance, the skeptic's claim is refuted, for, unchallenged in any specific way, common sense reasserts itself.⁷

Ouine wants to show that the ordinary view of translation as expressing the same meanings in different languages involves a "scientific mistake" akin to believing in the gods of Homer. To show this, he sets out to show that the intensionalist tradition from Kant to Carnap which claims to make objective sense of this view is, at bottom, no better than mythological explanation. Quine goes right to the heart of the matter: the relation of translation. Translation is critical because it is the only relation that provides interlinguistic identity conditions that are discriminating enough to individuate the fine-grained propositions of intensionalism. Such maximally finegrained propositions are what enable intensionalists to claim that their position is the only one that does justice to the full range of our ordinary, pretheoretical intuitions about linguistic structure. For example, only conditions for propositional identity based on synonymy seem capable of accounting for the intuition that the sentences "The sentence 'Two is less than three' means that two is less than three' and "The sentence 'The even prime is less than three' means two is less than three" have different truth values. Also, Frege, as is well known, argued (in "On Sense and Reference") that reference to such maximally fine-grained propositions is necessary in order to formulate the principles of logic. Hence, without the relation of translation, the intensionalist can claim no advantage in the study of language, and Frege's move to take propositions in logic as senses of sentences in language does not get off the ground. Thus, if Quine can establish that no objective sense can be made of equivalence of meaning for sentences of natural languages, intensionalism will be discredited as completely as the Homeric creation myths.8

Some philosophers offer a quick rebuttal to this line of argument. They claim that talk about meaning no more requires a statement of identity conditions to legitimatize it than talk about such things as nations or works of art require a statement of identity conditions to legitimatize them. But neither Frege nor other intensionalists with the hope of vindicating intensionalist semantics as the best scientific account of language can afford so cavalier an attitude. Although, in the early stages of research, it might be necessary for intensionalists

⁷ It is, of course, one thing having the common-sense view of translation and quite another having the kind of justification for it that would be provided by a full-fledged linguistic theory of translation. Although having both is clearly more desirable than having only the former, the latter is not necessary to refute the skeptic's claim.

⁸ Philosophy of Logic (Englewood Cliffs, NJ: Prentice-Hall, 1970), p. 3.

to tolerate entities without identity, it would not be advisable for them to build such indulgence into the research plan. Hence, I shall accept the need for identity conditions on exactly Quine's terms. If he can show that translation is indeterminate, I shall straight off concede that sense make no sense.

The character of Quine's argument for indeterminacy is indicated in such passages as:

. . . if the posit of propositions is to be taken seriously, eternal sentences of other languages must be supposed to mean propositions too; and each of these must be identical with or distinct from each proposition meant by an eternal sentence of our own . . . Surely it is philosophically unsatisfactory for such questions of identity to arise as recognized questions, however academic, without there being in principle some suggestion of how to construe them in terms of domestic and foreign dispositions to verbal behavior (WO, p. 205).

The aim of Quine's argument is thus to establish that, in principle, there is no way to construe questions about identity of intensional objects in terms of objective facts about verbal behavior. The reason is that translation is indeterminate in the sense that

. . . manuals for translating one language into another can be set up in divergent ways, all compatible with the totality of speech dispositions, yet incompatible with one another. In countless places they will diverge in giving, as their respective translations of a sentence of one language, sentences of the other language which stand to each other in no plausible sort of equivalence however loose (*ibid.*, p. 27).

I shall make no objection to the issue being put in terms of speech dispositions. It seems to me much the same thing whether we talk of a speaker's speech dispositions, in particular, dispositions to characterize sentences verbally, or of a speaker's overtly expressed intuitive judgments about sentences. Hence, if there is no way to construe questions of identity of sense in terms of speech dispositions, I am prepared to concede that intensional objects are on a par with Homeric gods.

I shall also make no objection to Quine's statement that "the behaviorist approach is mandatory." The behaviorism he has in mind here is not the dreaded reductive doctrine of days gone by, but merely a way of putting the study of language on a par with other sciences by requiring the linguist's theoretical constructions to be justified on the basis of objective evidence in the form of the overt behavior of speakers. Quine writes:

⁹ "Indeterminacy of Translation Again," this JOURNAL, LXXXIV, 1 (Jan. 1987): 5-10.

In psychology one may or may not be a behaviorist, but in linguistics one has no choice. Each of us learns his language by observing other people's verbal behavior and having his own faltering verbal behavior observed and reinforced or corrected by others. . . . There is nothing in linguistic meaning, then, beyond what is to be gleaned from overt behavior in overt circumstances (*ibid.*, p. 5).

Quine's behaviorism is thus a behaviorism one can live with. (Indeed, if it were not, it would be too controversial in the present cognitive climate to bear the weight of the indeterminacy argument.) Quine's behaviorism merely takes linguists out of their arm chairs and puts them in the field facing the task of having to arrive at a theory of a language on the basis of the overt behavior of its speakers in overt circumstances. Instead of challenging Quine's behaviorism, I shall challenge his claim about what can be gleaned from such behavior in such circumstances. I shall try to show there is intensional grain to be gathered.¹⁰

Quine's conclusion that the totality of linguistic evidence cannot eliminate incompatible translation manuals is developed in the situation he calls "radical translation." He illustrates this situation with a jungle story about a field linguist trying to choose among putative translations for the expression 'gavagai' in an alien language. Quine argues that the informant's dispositions to verbal response in such a translation situation are "incapable of deciding among 'rabbit', 'rabbit, stage', and various other terms as translations of 'gavagai' " (WO, pp. 71/2). Although nothing near a proof, the argument exhibits an unbreakable symmetry among the evidential considerations that can be adduced to justify the various translations. Whatever we can say on behalf of one translation, we can also say on behalf of the others. The reason is that the ostensive acts of the field linguist and the informant cannot refer to a rabbit without referring to a rabbit stage or an undetached rabbit part, nor any of these without the others, and radical translation contains nothing that enables the linguist to impose controls on hypotheses which enable them to choose between extensionally equivalent translation options. Thus, the argument leaves us with no grounds on which to resist its con-

Quine argues that reflection on "the nature of possible data and method" in radical translation suffices to make us "appreciate the indeterminacy":

Why does Quine not think behaviorism is mandatory for the psychologist as well? Does he not suppose that we learn more than just languages by observing others and being corrected by them? Is "introspective semantics" (*ibid.*, p. 9) no good, but introspective psychology alright?

There can be no doubt that rival systems of analytical hypotheses can fit the totality of speech behavior to perfection, and can fit the totality of dispositions to speech behavior as well, and still specify mutually incompatible translations of countless sentences insusceptible of independent controls (WO, p. 72).

But, despite the assurance, there can be doubts. I concede that there is no doubt about Quine's conclusion if restricted to radical translation. But it is not clear how actual translation, to which Quine's conclusion must also apply if indeterminacy is to matter philosophically, is related to radical translation. Thus, I think there can be doubts about the step in Quine's argument from his account of radical translation to his conclusion as applied to actual translation.

The doubt may be fleshed out as follows. Radical translation, it can be argued, is Quine's creation. He constructed it so that nothing beyond referential considerations provide evidence for translations. Thus, it is guaranteed that any evidential support for one among a set of referentially indiscernible hypotheses, such as "rabbit," "rabbit stage," and "undetached rabbit part," can be matched with equal evidential support for each of the others. But could not Quine's creation fail to mirror relevant evidential features of actual translation situations? We have, as yet, no grounds for rejecting the claim of intensionalists that actual translation situations provide linguists with evidence relevant to sense differences among the competing hypotheses which can be used to discriminate among them. If linguists in actual translation situations have such "independent controls," there is no evidential symmetry in those situations, and, hence, no indeterminacy.

My point so far is only that Quine has to say something to make us "appreciate the indeterminacy" for actual translation. Quine needs to clarify the relation between radical translation, as he defines it, and actual translation, as it exists or could exist in the practice of real linguists. Quine seems to address this need. He presents radical translation as the limiting case of actual translation, i.e., as the case where historical differences between the languages and cultural differences between its speakers are maximal. It is presented as the most philosophically perspicacious case of actual translation in virtue of being the one where the issue about meaning is least likely to be confused by historical and cultural similarities.

But the issue is not so easily settled. We have been given likenesses between radical translation and the limiting case of actual translation, but have not as yet been given a reason for identifying them.¹¹

¹¹ Note, for example, how in "Indeterminacy of Translation Again" Quine slides unfalteringly from speaking of "translation" and "our linguist" (pp. 5/6) to speaking of "radical translation" and "our radical translator" (p. 5/6) to speaking of "radical translation" and "our radical translator" (p. 47) (pp. 5/6) to speaking of "radical translation" (pp. 5/6) to speaking of "radical" (pp. 5/6) to speaking of "radica

The acceptability of an identification depends on whether actual translation is in all relevant respects like radical translation. Now, one relevant respect is surely whether or not, in actual translation, the matching of expressions as synonymous also takes place in the absence of "independent controls." For it is the absence of such controls in radical translation which causes evidential symmetry and indeterminacy. Hence, we must be given a reason for believing such controls do not exist in actual translation.

Quine has what he thinks is a reason: the existence of such controls depends on the existence of intensional objects in the way that the existence of independent controls in physics depends on the existence of physical objects; but to suppose that "translational synonymy at its worst is no worse off than physics" is to "misjudge the parallel" (WO, p. 75). There is a fundamental difference between intensional semantics and a genuine science like physics: in physics, "the parameters of truth stay conveniently fixed most of the time; not so with "the analytic hypotheses that constitute the parameter of translation" (WO, p. 76). Quine explains:

Something of the true situation verges on visibility when the sentences concerned are extremely theoretical. Thus who would undertake to translate 'Neutrinos lack mass' into jungle language? If anyone does, we may expect him to coin words or distort the usage of old ones. We may expect him to plead in extenuation that the natives lack the requisite concepts; also that they know too little physics. And he is right, except for the hint of there being some free-floating, linguistically neutral meaning which we capture in 'Neutrinos lack mass', and the native cannot (WO, p. 76).

There is no domain of linguistically neutral meanings corresponding to the domain of physical objects, and, consequently, no facts against which to judge the truth of analytical hypotheses, since such hypotheses assert that a sentence in the target language expresses the same linguistically neutral meaning as one in the home language. Thus, if there are no meanings, it makes no sense to talk of a scientific choice between competing analytical hypotheses.

As Quine sees it, the correct comparison with physics is this. Theories in physics are underdetermined by the available observational evidence and also by the total possible evidence, but not subject to indeterminacy (WO, pp. 75/6). Underdetermination is only a matter of "empirical slack" which can be taken up methodologically, that is, genuinely divergent physical theories which survive confrontation with the total evidence can be adjudicated by appeal to

¹² Up-dated in "On the Reasons for Indeterminacy of Translation," this JOUR-NAL, LXVII, 6 (March 26, 1970): 178–183.

methodological canons like simplicity, depth of explanation, etc. The differences between such theories are substantive because there is a physical fact they are about. Intensional semantics, in contrast, suffers from a condition far worse than underdetermination. It suffers from indeterminacy whose etiology is the lack of a fact of the matter for semantic theories to be right or wrong about (WO, pp. 76–79; also LPV, p. 63). Therefore, whereas the condition of physics can be treated methodologically, the illness of intensional semantics is terminal.

But this reason for thinking independent controls do not exist in translation is only as good as the reason Quine has for saying that there are no linguistically neutral meanings. It is not enough for Ouine just to claim that "[t]he discontinuity of radical translation tries our meanings: really sets them over against their verbal embodiments or, more typically, finds nothing there" (WO, p. 76). Nothing is established by a mere claim. Moreover, if Quine is employing "the discontinuity of radical translation" to argue for there being no fact of the matter in actual translation, then, once the question is put as we have put it, i.e., in terms of how the step from radical translation to actual translation is justified, this employment seems to beg the question. At this point, all that Quine can legitimately say is that there is no fact of the matter about meaning in the translation situation that he invented. He is not entitled to say that there are no linguistically neutral meanings in actual translation. Recall that our earlier doubts about the identification of radical translation with the extreme of actual translation arose because Quine had not established that the cases do not relevantly differ with respect to the existence of "independent controls." But the existence of independent controls, as we have seen, is not unrelated to the existence of meanings.

Furthermore, even if Quine had secured the step from radical translation to actual translation, he still would not have a basis for claiming that there are no linguistically neutral meanings in actual translation. For, in fact, he has not even ruled out the possibility of meanings in radical translation! He has shown that there is no evidential basis for choosing between rival analytical hypotheses, but this establishes no more than the unknowability of meanings. But meanings, like Kant's noumena, could exist even if unknowable. No mere epistemological considerations, such as those in Quine's discussion of radical translation, entail an ontological conclusion such as he draws about linguistically neutral meanings. Such a conclusion introduces an ontological skepticism over and above his already asserted epistemological skepticism. Thus, rather than motivate the

claim that actual translation is indeterminate, the ontological skepticism only increases the burden of proof.

Hence, we have a new question: What is Quine's argument for claiming that there are no linguistically neutral meanings? 13 I think, without doubt, that he has one. It is a mistake for Noam Chomsky to represent Quine as simply stipulating that linguistics, in contrast to other sciences, can have no general theories. 14 It is true that Quine does not at this point explicitly present an argument, or even cite one, but surely a philosopher as acute as Quine must see that an argument is needed to back up his claim. He must know that, without one, indeterminacy of translation is unsupported, and intensional semantics can be accused of nothing more serious than underdetermination. In this case, the omnipresence of divergent translations is of no philosophical interest, indicating nothing more exciting than gaps in knowledge of semantic fact or insufficient applications of scientific methodology.

It would also be a mistake to suppose Quine is simply appealing to some well-known philosophical doctrine like behaviorism to back up his claim that there are no semantic facts. As we have seen, his behaviorism would not be equal to the task. It is not the militant doctrine which brands as scientific heresy everything that cannot be strictly defined in terms of stimulus and response. Furthermore, Quine's claim does not derive from old-fashioned verificationism. If Quine has verificationist scruples, they have to be rather mild ones in order to allow him to countenance highly theoretical entities in science, such as the ten dimensional wonders of contemporary physics and the objects of set theory. Finally, the claim could not be a consequence of physicalism, either. Physicalism would allow meanings so long as they are reducible to brain states. Quine's view is not that intensionalists have been lax in showing that meanings are reducible to physical states, but that they have been deluded in thinking there are any such things to be reduced. It would be absurd for a physicalist to undertake a physicalistic reduction of Homeric gods.

We get a clue to what Quine's argument really is by noting that he makes the claim that there are no linguistically neutral meanings with the confidence of someone introducing a lemma he has already

"Quine's Empirical Assumptions," in Words and Objections, Donald Davidson and Jaako Hintikka, eds. (Boston: Reidel, 1969), p. 62.

We need not interpret Quine as taking the discontinuity of radical translation to be a basis for inferring the absence of linguistically neutral meanings. Perhaps all he means is that such discontinuity helps us see that there are not any meanings. But, even on this interpretation, the question in the text stands, since, here too, without a justification, Quine's case would come down to a bare assertion.

proved and can reasonably expect his readers to know. I submit that this is exactly it: Quine thinks he has already given a conclusive argument against meanings in his earlier works, particularly in his famous paper "Two Dogmas of Empiricism." Indeed, the argument there is directly to the point. It aims to show that we can make no objective sense of synonymy, and, if this argument were conclusive, no respectable theory would quantify over meanings and considerations of parsimony would oblige us to deny there are meanings.

That it is correct to interpret Quine's argument for indeterminacy to have this reference back to his earlier work is shown not only by the tone and logic of his reasoning in Word and Object, but also by explicit statements about his overall anti-intensionalist strategy in various places. In the early paper "The Problem of Meaning in Linguistics" (LPV, pp. 47-64), Quine makes it clear that, even then, he thought of his criticism of the analytic-synthetic distinction as undercutting the basis for claims that there is a fact of the matter in connection with meaning. Referring to a restatement of his arguments against synonymy from "Two Dogmas of Empiricism," Quine says that construction of a lexicon for translating a language from a radically different culture is not a well-defined task because such construction suffers from a "paucity of explicit controls" (LPV, p. 63). (Note we even have the same notion of absence of controls which figures so prominently in Word and Object.) Quine completes the thought, saying:

The finished lexicon is a case, evidently, of ex pede Herculem. But there is a difference. In projecting Hercules from the foot we risk error, but we may derive comfort from the fact that there is something to be wrong about. In the case of the lexicon, pending some definition of synonymy, we have no statement of the problem; we have nothing for the lexicographer to be right or wrong about (LPV, p. 63).

So, if a definition of synonymy is ruled out by "Two Dogmas of Empiricism," then there is nothing for the lexicographer to study.

In the recent paper "Indeterminacy of Translation Again," Quine provides further evidence for taking "Two Dogmas of Empiricism" to provide the argument for his claim in Word and Object that there are no meanings:

Considerations of the sort we have been surveying are all that the radical translator has to go on. This is not because the meanings of sentences are elusive or inscrutable; it is because there is nothing to them, beyond what these fumbling procedures can come up with. Nor is there hope even of codifying these procedures and then defining what counts as translation by citing the procedures; for the procedures involve weighing incom-

mensurable values. How much grotesqueness may we allow to the native's beliefs, for instance, in order to avoid how much grotesqueness in his grammar or semantics? (p. 8).

The point here is that no comparison is possible between hypotheses about beliefs and hypotheses about meaning, because such a comparison assumes an analytic-synthetic distinction. If there were such a distinction, the linguist could, in principle, decide whether a piece of information belongs in the theory of the informant's language or in the theory of his or her extra-linguistic beliefs. Without an analytic-synthetic distinction, such decisions involve "weighing incommensurable values."

Given that "Two Dogmas of Empiricism" is intended to supply the argument for the claim that there are no meanings which does not explicitly appear in Word and Object, Quine's overall argument for indeterminacy of translation can be reconstructed as follows. Assuming that "Two Dogmas of Empiricism" removes the possibility of a linguistically universal synonymy relation,15 there can be no identity conditions for intensional objects, and, as a consequence, we must abandon the idea of linguistically neutral meanings serving as the common content of a sentence and its translation. Thus, there is no parallel between semantics as conceived in traditional intensionalism and bona fide sciences like physics. In physics, there are objects of study, and so physics suffers only from underdetermination. In semantics, there are no objects of study, and, hence, there can be no evidence to provide controls on analytical hypotheses and to make objective sense of talk about rational choices among theories of meaning. Thus, there is nothing to distinguish actual translation from radical translation, and Quine can identify the latter with the limit case of the former. The symmetry argument for radical translation transfers to actual translation, and he can conclude that, in actual translation, divergent systems of analytical hypotheses fit the totality of speech dispositions to perfection. He can then say, justifi-

¹⁵ Another line of argument is this. Quine's aim is to show, contra Alonso Church, Frege, and Carnap, that there are no intensional objects for intensionalists to use in their statement of principles of logic. Because it is into logic that intensionalists propose to introduce senses, the relation of synonymy which provides the identity condition must be specified, as Quine puts it in "Two Dogmas of Empiricism" (pp. 33/4), for variable 'S' and 'L'. The reason is that principles of logic express a language neutral implication relation. That is to say, logic concerns a notion of implication which is specified for variable 'S' and 'L', not a notion like implication-in-Italian. Thus, if Quine can show there is no linguistically universal notion of synonymy, there will be no linguistically neutral meanings, and, hence, nothing for intensionalists to appeal to extend logic in a naturalistically dubious direction.

ably, that the limits of possible data for radical translation make the indeterminacy of translation certain.¹⁶

Having located the "missing" argument, I now want to show that it does not work, and, hence, that the argument for indeterminacy does not work either. The argument in "Two Dogmas of Empiricism" takes the form of a proof by cases. It begins with an enumeration of the areas where it would be reasonable to look for an explanatory paradigm to use in trying to make objective sense of the concepts in the theory of meaning. The areas are definition generally, logical theory, and linguistics. Quine asks whether the methods for explaining concepts in any of these areas can explain synonymy and analyticity. He examines the areas in turn. He argues that the available paradigms in the case of definition either assume prior synonymy relations or else have nothing to do with meaning (LPV, pp. 24-27). In the case of logical theory, he argues that Carnapian meaning postulates and semantical rules shed no light whatever on the nature of synonymy and analyticity (LPV, pp. 32-37). In the case of linguistics, he argues that the methods for defining concepts are demonstrably unable to provide noncircular definitions of these concepts (LPV, pp. 27-32). Since these cases exhaust the areas where we might expect to find an explanatory paradigm appropriate to logico-linguistic concepts like analyticity and synonymy, Quine concludes that there are no methods for clarifying synonymy and analyticity.

Quine's argument in the case of definition is absolutely compelling. So is his argument in the case of logic. But the final argument needed to complete the proof by cases, the argument for the case of linguistics, is quite a different matter. Once its structure is revealed, it will be clear that it does nothing to establish that attempts to explain analyticity and synonymy in linguistics must fail. I shall discuss this argument in some detail both because of its importance to the indeterminacy thesis and because the unfamiliarity of philosophers with linguistics is surely a major factor in the readiness with which they have accepted Quine's account of its methods.

Quine's argument begins by identifying *substitution criteria* as the proper method in linguistics for defining concepts like analyticity and synonymy. He explains as follows:

So-called substitution criteria, or conditions of interchangeability, have in one form or another played central roles in modern grammar. For the synonymy problem of semantics such an approach seems more obvious still. However, the notion of the interchangeability of two linguistic

^{16 &}quot;Indeterminacy of Translation Again," p. 9.

forms makes sense only in so far as answers are provided to these two questions: (a) In just what sorts of contextual positions, if not in all, are the two forms to be interchangeable? (b) The forms are to be interchangeable salvo quo? Supplanting one form by another in any context changes something, namely form, at least; and (b) asks what feature the interchange is to leave invariant. Alternative answers to (a) and (b) give alternative notions of interchangeability, some suited to defining grammatical correspondences and others, conceivably, to defining synonymy (LPV, p. 56).

It is important to recognize that the substitution criteria Quine borrows from "modern grammar" are not simply the customary substitution operations in logic and mathematics. To be sure, like those operations, substitution criteria specify a concept on the basis of a feature which remains invariant when and only when the elements that replace one another in the chosen context belong to the extension of the concept. But, in the case of the substitution criteria from "modern grammar," there is the special requirement that statements of the context and of the invariant feature not contain concepts belonging to the family of the concept to be defined. Thus, definitions taking the form of substitution criteria may fail either because the feature does not correlate with all and only expressions in the extension of the concept or because the special requirement is not met, that is, the feature or the substitution context is stated using the concept to be defined or a concept in its family.

Given that substitution criteria are the proper method for defining linguistic concepts, Quine has an easy time demonstrating that the concepts of analyticity and synonymy cannot be defined in linguistics. Suppose we wish to define synonymy. The context must be either intensional or extensional. If we choose an intensional context, say "Necessarily, ______," we can use truth as the feature which is to remain invariant in substitutions, but then we violate the requirement of noncircularity at the outset because the context has to be characterized using synonymy or other concepts from the theory of meaning. If we choose an extensional context, we get the definition off the ground, but then we can no longer use truth as the feature which is to remain invariant. For, in extensional contexts, truth will not discriminate coextensive but nonsynonymous expressions from synonymous expressions. It is thus necessary to move to something stronger like necessary truth or analyticity. But this vio-

¹⁷ Quine observes that Carnap's possible worlds account of necessity uses meaning relations among state descriptions to guarantee logical independence (LPV, p. 23).

lates the noncircularity requirement, too, because such notions are

defined in terms of synonymy.

This, then, is the critical argument on which the argument for indeterminacy in *Word and Object* depends. The trouble with it is that the assumption from which it proceeds, that substitution criteria are the proper way to clarify concepts in linguistics, is easily knocked down. There is nothing in favor of the assumption, and there are strong *a priori* and historical arguments against it.

Quine does not justify relying on substitution criteria. Rather, he notes that this approach to definition has "in one form or another played central roles in modern grammar." This, however, is only to say that the approach occupied a central place in the phonological and syntactic investigations which took place during the Bloomfieldian period in linguistics (roughly 1930s-1950s). Quine offers nothing beyond their place in Bloomfieldian linguistics to show that substitution criteria are indispensible to the science of language and hence a valid approach to defining synonymy. But the fact that one school of thought in a science at one time in its history has a particular methodological practice means very little, given how frequently we see schools of thought come and go in science and old paradigms replaced with new ones. In fact, even as Quine wrote, the positivist foundations of Bloomfieldian linguistics were being eroded by what we can now see was a successful critique of logical positivism in the philosophy of science.18

An a priori reason for thinking that substitution criteria are neither the only nor the preferred form of definition in linguistics is that physics, mathematics, and logic provide an example of another form of definition which could be adapted for concepts in linguistics. This is the familiar approach of defining a concept on the basis of an axiomatic or recursive specification of the relations between it and other concepts in its family, as, for example, in the Dedekind-Peano axiomatization of arithmetic concepts, or the axiomatization of the truth-functional connectives in a standard sentential calculus. The difference between such theoretical definitions, as I shall call them, and substitution criteria is that theoretical definitions can use concepts belonging to the same family as the concept to be defined. Effective theoretical definitions explain such concepts in a way which captures the structure of the primitives in the family of concepts. The degree of relatedness exhibited among the concepts in the fam-

¹⁸ To see how important a role positivism played in shaping Leonard Bloomfield's reconstruction of linguistics, see his "Language or Ideas?" *Language*, XII (1936): 89–95; reprinted in *The Philosophy of Linguistics*, J. J. Katz, ed. (New York: Oxford, 1985), pp. 19–25.

ily is thus a measure, not of circularity, but of the systematizing power of the explanation.

This difference emerges sharply when we try to impose a substitution criteria definition on subjects like logic and mathematics. Imitating Quine's demand that semantic concepts be defined by substitution criteria, we could demand that truth-functional connectives and numbers be defined by them. If this were a legitimate demand, we would be able to construct an argument directly parallel to Ouine's showing that sentential logic and arithmetic make no objective sense. For, as with synonymy, there is no noncircular property which is invariant on all and only substitutions of logically equivalent propositions, on the one hand, and substitutions of numerically identical quantities, on the other. Now, such an argument is clearly a reductio of the idea that legitimacy of logic and mathematics depends on their identity relation being definable by means of substitution criteria. Thus, it is open to intensionalists to say that Quine's own argument is a reductio of the idea that the legitimacy of the theory of meaning depends on its identity relation being definable by means of such criteria. The circularity that Quine exhibited can thus be seen as a product of imposing an inappropriate definitional paradigm on the theory of meaning.

These a priori considerations show that there is an alternative definitional paradigm possible for linguistics which, in certain ways, is more promising than substitution criteria. It does not matter, given the logic of my argument, whether such an alternative ever actually finds its way into linguistics. It could be introduced at any time as another, and preferable, way of defining the family of semantic concepts. What could prevent it? Surely not Quine's scruples about "misjudging the parallel." These scruples can have force, as we have seen, only after we are given a persuasive argument against linguistically neutral meanings, for only then is there a relevant distinction between intensional semantics and genuine science. At this stage, however, Quine has yet to put such an argument together.

Furthermore, it would do no good to appeal to the arguments Quine used to show that Carnapian meaning postulates and semantical rules fail to explain analyticity and synonymy (LPV, pp. 32–36). Systems of meaning postulates and semantical rules are one way of constructing a theoretical definition for semantic notions, but not the only way. The fact that one way fails to explain them does not show that others will fail. Moreover, Carnapian systems have an idiosyncratic feature which is what makes them subject to Quine's criticism that they are unilluminating about analyticity and synonymy, namely, they do not concern these sense concepts, but the

Digitized by Arya Samaj Foundation Chennai and eGangotri

broader concepts of necessary truth and necessary equivalence. The postulates of such systems are not about the senses of words but about their referents. They do not describe the structure of senses, in the way phrase markers describe the syntactic structure, but, like the logical postulates they are modeled on, they just constrain the extensional interpretations of a language. Ironically, so-called meaning postulates are not about meaning.¹⁹

There is also a historical reason for thinking substitution criteria do not deserve the place Quine accords them. Quine supposed that the methodology of Bloomfieldian linguistics is an indispensible aspect of the science of language, and, hence, something that he could rely on in philosophy. This supposition was soon challenged. In the early 1960s, the field of linguistics underwent what has come to be called "the Chomskian revolution." One of the principal changes which the revolution brought about was a paradigm shift from substitution-criteria definition in taxonomic grammars to theoretical definition in generative grammars.20 Chomsky explicitly modeled his conception of a generative grammar on formal systems in logic. The theorems, instead of being a class of logical truths, are the well-formed sentences of a language. An optimal generative grammar for a language L generates all and only the well-formed sentences of L. It is thus a recursive definition of the notion "sentence of L." The derivations of a sentence provide a description of its grammatical structure.

Chomsky carried the analogy with logic further. He modeled his conception of general linguistics (i.e., the study of linguistic universals) on metalogic. The definitions of language neutral concepts were to be given in terms of theoretical definitions stated in the metatheory for generative grammars (i.e., linguistic theory, in Chomsky's terminology). For example, Chomsky defined the concept of a syn-

²⁰ One of the most influential examples of theoretical definition in the early stages of generative syntax was Chomsky's rules for the English auxiliary system; see Current Issues in Linguistic Theory (The Hague: Mouton, 1964), p. 36.

¹⁹ It is important to note that Carnap's work represents a significant break with the Fregean tradition preceding it. Whereas that tradition characterized analyticity in terms of intensional notions (e.g., Frege characterized it in terms of laws of logic plus definitions, i.e., statements of sameness of sense), Carnap abandons the use of intensional notions in his meaning postulates approach, characterizing analyticity for failing to explain what analyticity is. If analyticity can be explained, you have to to other intensionalist accounts of analyticity. See my Cogitations (New York: Oxford, 1986), pp. 41–97.

tactically well-formed sentence of a language as a sentence whose syntactic representation can be generated in an optimal grammar of the language, and the concept of two constituents being of the same syntactic type as identity of their syntactic representations in an optimal grammar.21

In the context of the present discussion of Quine, it is hard to exaggerate the importance of the shift to the new paradigm of generative grammar. The new paradigm, together with the theoretical definitions worked out for concepts in phonology and syntax, provided linguists and philosophers with a model for theoretical definitions in semantics (just as substitution criteria in phonology and syntax had provided Quine a model). Of special significance in this connection is the possibility opened up by the metatheory for generative grammars of giving language-independent definitions of language-neutral concepts. This enables us to avoid the use of language-specific definitions such as meaning postulates, and, thus, as will be explained below, to escape Quine's criticism of such definitions. Theoretical definitions in linguistic theory provide a way of defining concepts in the theory of meaning for variable 'S' and 'L' because, in defining a concept at the level of linguistic theory, they define it in terms of features of optimal generative grammars for every natural language.

Research in semantics within the generative framework during the 1960s and early 1970s exploited the model of theoretical definitions in generative phonology, generative syntax, and linguistic theory.22 The aim of this research was to define concepts such as meaningfulness, ambiguity, synonymy, analyticity, etc., on the same basis as syntactic concepts like well-formedness. Constructing such definitions involved two steps. First, it is necessary to develop a conception of the formal representation of sense structure which parallels the formal representation of constituent structure provided in syntactic markers. But, instead of describing the way sentences are built up from constituents like nouns, verbs, prepositions, etc., as syntactic markers do, semantic markers have to describe the way senses of sentences are built up from the senses of their syntactic constituents.

²¹ Syntactic Structures (The Hague: Mouton, 1957), pp. 49-60, and especially

pp. 53-55 for discussion of theoretical definitions.

See my Semantic Theory (New York: Harper & Row, 1972); Propositional in Semantics. Frank LePore ed. (London: in Semantics," in New Directions in Semantics, Ernest LePore, ed. (London: Academic, 1987), pp. 157-233.

Second, it is necessary to find the formal features of semantic representations which, from sentence to sentence and language to language

guage, correlate with particular semantic concepts.

Much of this research was focused on the concept of analyticity. A type of semantic representation was developed which makes it possible formally to describe analytic sense structure, namely, decompositional semantic representation. ²³ The symbols in such representations stand for component senses of the sense being represented and for their relations to one another. A decompositional representation of "bachelor," for example, would contain symbols for the component senses 'human', 'adult', 'male', and 'single'. Decompositional representation is the only type that can explicate the Kantian concept of analyticity, that is, judgments which add

. . . nothing through the predicate to the concept of the subject, but merely break . . . it up into those constituent concepts that have all along been thought in it. 24

This is because it is the only type of semantic representation which marks the presence of the sense of the predicate in the sense of the subject in sentences like 'Bachelors are male' where both terms are syntactically simple.

Relative to an assignment of decompositional representations to sentences, we can define analyticity in terms of the semantic representations of the full predicate and each of its terms but one being formally contained in the semantic representation of the remaining term. Similarly, we can define meaningfulness and synonymy. Semantic well-formedness can be defined in terms of the generability of at least one semantic representation for an expression. Semantic identity can be defined in terms of sameness of semantic representation for two expressions.

The first thing to note about such a definition of analyticity is that it makes no reference to thought processes, and, accordingly, avoids Frege's criticisms of Kant's psychologism. The second thing to note is that the definition is broader than Kant's, which is restricted to

²⁴ Kant, The Critique of Pure Reason, Norman Kemp Smith, trans. (New York: Humanities, 1929), p. 49

²³ Cogitations, pp. 75-90.

²⁵ This is but the sketchiest of presentations, but see the references in footnote 22. A full presentation is found in G. E. Smith and J. J. Katz, *Supposable Worlds* (Cambridge: Harvard, forthcoming).

subject-predicate sentences. Our definition covers non-subject-predicate sentences like 'Smith marries those he weds' and 'Jones buys books from those who sell her books'. This feature avoids another of Frege's criticisms of Kant. Finally, note that the definition avoids both of Quine's criticisms of Carnap's explication. As already indicated, the definition, being in the metatheory of generative grammars, meets Quine's demand that the concept of analyticity be defined for variable 'S' and 'L'. 26 It also meets Quine's demand that a definition tell us what property is attributed to a sentence when marked "analytic." The definition says the property is that of having a redundant predication—the referential upshot of which is that the truth conditions of the sentence are automatically satisfied once its terms take on reference.

Although more has been accomplished than the above sketch indicates, nothing even approaching a full theory of decompositional semantics has been developed. The fact that such a theory is still far from complete is, however, not something that Quine can use to block the criticism I am making of the argument in "Two Dogmas of Empiricism." In particular, it is not open to him to argue that acceptance of the criticism must be withheld until the theory of decompositional semantics is developed far enough to see whether it works. Since Quine's argument claims to rule out every possibility of showing that intensional concepts can be made sense of, showing that his argument overlooks a possibility is showing that it fails.²⁷

I now wish to trace out the consequences of Quine's failure to support his claim that there are no linguistically neutral meanings. The immediate consequence is that we are free to entertain the existence of meanings in the same spirit with which scientists beginning the study of a new field presume the existence of facts and laws

²⁷ Moreover, theoretical definitions of semantic structure occupy a natural place in the theory of generative grammar, extending that theory by adding a theory of semantic structure to the theories of phonological and syntactic structure. Further, as we indicated, the use of theoretical definition is what makes Quine's criticism of vicious circularity inapplicable to fields like logic and mathematics.

Quine has somehow not seen the point that a theoretical definition of analyticity framed in the metatheory of generative grammars defines the notion for variable 'S' and 'L', which I first made in my "Some Remarks on Quine on Analyticity," this JOURNAL, LXIV, 1 (Jan. 19, 1967): 36–52. See Quine, "On a Suggestion of Katz," ibid., pp. 52–54; and also "Methodological Reflections on Current Linguistic Theory," in Semantics of Natural Language, Donald Davidson and Gilbert Harman, eds. (Dordrecht: Reidel, 1972), pp. 449–450. In these papers, he confuses a theoretical definition of analyticity with a technique for questioning informants to elicit judgments about analyticity for the sentences of a particular natural language.

to be discovered. Without a reason to think that the theory of meaning suffers from anything worse than underdetermination, we may pursue the parallel with well-developed sciences. We may take the view that such a course of action is itself the best way to determine whether there are meanings. As with other sciences, trying to construct a theory will, in the long run, show whether the initial presumption is correct.

As a consequence of thus being able to pursue the parallel, Quine's argument that actual translation is indeterminate breaks down. Recall that this argument depends critically on whether actual translation, like radical translation, lacks "independent controls." As Quine claims: "when we reflect on the limits of possible data for radical translation, the indeterminacy is not to be doubted."28 We conceded that the constraints on the choice of analytical hypotheses in radical translation are too weak to make translation determinate. but questioned why the constraints in actual translation are supposed to be no stronger. Pressing this question, we found nothing to stop us from making the working assumption that there are linguistically neutral meanings. But, if we can assume a fact of the matter with respect to which analytical hypotheses can be judged, we can say what constraints there are in actual translation over and above those in radical translation. This, in turn, will explain how we find out whether the assumption is true.

The constraints on choices between analytical hypotheses are directly parallel to those we use to make choices among corresponding hypotheses in other sciences. The criteria for correctness in translation will be the customary blend of data and methodological considerations (i.e., the data must be explained as economically as possible, with as much scope and depth as possible, etc.). The data come from overtly expressed judgments of speakers reflecting their knowledge of the language. In radical translation, the data are restricted to judgments about the reference of expressions. But, on our working assumption, the data in actual translation include judgments about the senses of expressions, too. The linguist can ask whether an expression has a sense (i.e., whether it is meaningful or not), whether an expression has more than one sense (i.e., whether it is ambiguous), whether the sense of one expression is the same as that of another (i.e., whether they are synonymous), whether the sense of one expression is the opposite of the sense of another (i.e., whether

²⁸ "Indeterminacy of Translation Again," p. 9.

they are antonyms), whether the sense of a sentence involves redundant predication (i.e., whether it is analytic), and so on. The possibility of putting such questions to informants automatically provides the possibility of evidential controls. Such controls take the form of the requirement to prefer those hypotheses which provide the best account for the data about sense structure, where the choice among accounts which handle the same range of data is made on the basis of standard methodological considerations.²⁹

To give some concrete examples of semantic evidence, I shall tell a jungle story of my own. Imagine an actual translation situation in which the linguist and informant come from disparate cultures and in which the home and target languages have disparate histories. The linguist is faced with the choice of translating 'gavagai' as "rabbit," "rabbit stage," or "undetached rabbit part." In my story, unlike in Quine's, the linguist simply queries the informant about the senses of expressions. (We shall consider the question of how they communicate below.) The linguist can ask whether "gavagai" is synonymous with one of these English words. The informant may judge it to be synonymous with one of them, judge it not to be synonymous with any, or provide no useful judgment. In the first two cases, the linguist has acquired some data. In the third, the direct approach did not work. But here, as elsewhere in science, there is a wide range of less direct approaches. For example, the linguist may ask whether "gavagai" bears the same relation to a native expression that "finger" bears to "hand" or "handle" bears to "knife." Or, the linguist could ask whether "gavagai" is closer in sense to "infancy" and "adolescence" than it is to "infant" and "adolescent." Or, the linguist could ask whether the expression obtained when "gavagai" is modified by the native word meaning "undetached" is redundant like "unmarried bachelor." Or, whether the expression obtained when "gavagai" is modified by the native word meaning "detached" is contradictory like "married bachelor." There are indefinitely many further questions to ask, and, of course, indefinitely many further informants to ask them of. So why should a body of data obtained in this way, in principle, not settle any question of translation as satisfactorily as a comparable body of data in other sciences would settle comparable theoretical questions?

Two Quinean thoughts arise here. First: there is the problem of

Note that our approach can adopt a holism which insists that translation schemes can only be tested (or even constructed) as components of a comprehensive grammatical theory of natural languages.

how linguists come by the hypotheses which they are supposed to verify by asking such questions about the foreign language. Second: there is the problem how linguists and informants are able to communicate.

It is hard to see why the first is much of a problem. The hypotheses can be based on guesses based on nothing, hunches educated by experience, prejudices stemming from a cherished theory, or what have you. This is surely how it is in other sciences. Yet Quine views presumptions about the semantics of an alien language deriving from the semantics of one's own language as begging the question.³⁰ He is surely right in the case of radical translation. If a linguist in such a situation takes "gavagai" to be an object term, like the English "rabbit," then, in the absence of "independent controls," there is no way to know whether the linguist has not merely read the grammatical categories of his or her own language into the alien language. But, for the case of actual translation, where, as we have shown, there can be such controls, presumptions about the semantics of the alien language, even those reflecting linguistic chauvinism, beg no question. Evidential controls enable linguists to revise mistaken presumptions and validate correct ones. Falsifications of the category structure of an alien language, accordingly, are on a par with the animistic or anthropomorphic theories of nature in early science. Presumptions on the part of the linguist may retard (or, for that matter, advance) the progress of translation, but, as long as evidence can be brought to bear on such presumptions, they have no philosophical significance.

The problem of communication between linguist and informant may at first seem more serious. Quine always represents radical translation situations as ones in which the linguist confronts the informant across an impassable language barrier. Indeed, Quine's assumption that they do not know each other's language has made his readers wonder how an informant knows what the linguist's interests are and how a linguist knows what behavior of the informant counts as assent. It is Quine's assumption of monolingualism—rather than his behaviorism or anything else—that is the truly unrealistic element in his thinking about translation. Actual translation can no more proceed without a bilingual than grammar construction can proceed without a native speaker.

³⁰ In WO, p. 72, Quine writes: "To project [analytical hypotheses] beyond the independently translatable sentences at all is in effect to impute our sense of linguistic analogy unverifiably to the native mind."

Digitized by Arya Samaj Foundation Chennai and eGangotri

There is, then, a sense in which the whole question of bilingualism is beside the point. Quine accepts the fact that indeterminacy arises equally well within a single language, between idiolects of different speakers, and, further, that it arises even within a single idiolect, between different temporal stages. Here there is no question of bilinguals, unless, of course, bilingualism is understood as fluency in two idiolects or two idiolect stages. Is Quine to be construed as claiming that English speaking linguists and their English speaking informants do not share a language in which they can communicate?

I am, of course, not saying that bilinguals must exist, but only that their existence is an empirical condition for translation. The existence of bilinguals is comparable to the existence of conditions that enable us to conduct the experiments necessary to decide among rival physical theories. Moreover, even if there were no bilinguals in the case of some alien language, we know a sure-fire method for creating them on demand. The method takes rather a long time, and its implementation involves various practical, social, and moral problems, but it works, as those who have acquired bilingual fluency growing up in a bilingual home can attest.

To be sure, Quine thinks introducing bilinguals begs the question. The feeling that a question is begged is strongest with respect to the possibility of full fluency in both languages, including, as it would, fluency in the metalinguistic vocabulary of the language, e.g., the expressions for relations like 'translates' or 'is synonymous with'. The idea of an informant who speaks the whole language like a native seems to go too far. But why? We are not supposing the actual existence of such informants, but only the possibility of their existence. Accepting the possibility of fluent informants (even their actuality) does not settle the question of the existence of meanings, anymore than accepting the possibility of reliable meters in a physics experiment (even their actuality) settles the question the experiment is to decide

Admittedly, entertaining the prospect of bilingualism is, in itself, entertaining the prospect of translation, since knowledge of translation is what makes a bilingual bilingual. But, if there is no harm in entertaining the prospect of translation, there is none in entertaining the prospect of bilingualism.

Thus, no question is begged. But, further, no risk is run. Although bilinguals, like field linguists with presumptions, can cause falsifications in translation, such falsifications, like false grammatical categories read into a target language, can be corrected. As long as we have evidential controls, the difficulties posed by deficient or devious

informants are, in principle, no different from those posed by a faulty meter in a physics experiment.

Let us sum up. Radical translation is indeterminate because it is restricted to using referential features of words. But there is nothing to show that actual translation is similarly restricted. With the use of evidence about the senses of words in constructing theories about actual translation, the evidential symmetry in radical translation does not arise in actual translation. Intensionalist semanticists may cherish the same hopes for success as other scientists. They are not even required to establish that facts about meaningfulness, ambiguity, synonymy, antonymy, redundancy, etc., are, in principle, sufficient to enable linguists to make a unique selection among a number of translation schemes. Either all but one of the translation schemes can be eliminated, given total evidence and methodological considerations, or else, since synonymy is the identity relation for meanings, the uneliminated schemes count equivalently as ways of expressing the truth about the semantic relation between the languages.

Of course, none of this is to say that no form of skepticism about determinate translation remains. There is no metaphysical insurance policy against nature being counterinductive or linguistic investigation being counterproductive: linguists can make systematically misleading projections, and informants can produce systematically misleading evidence. But such possibilities lead to nothing more than an absolute skepticism which would obliterate knowledge in all fields. Thus, once Quine's semantics relative skepticism goes, nothing prevents the common-sense view of translation from reasserting itself.

Life without meaning is not the trouble-free paradise it is sometimes made out to be. Quineans have always had trouble explaining in what respect allegedly competing analytical hypotheses compete. It is easy, from a common-sense standpoint, to appreciate the respect in which "rabbit," "rabbit stage," and "undetached rabbit part" represent rival translations, but what, from the standpoint of indeterminacy, is supposed to be the semantic difference? Quine's indeterminacy thesis is predicated on the existence of a conceptual distinction of some sort among such translations, yet the nature of the choice within a radical-translation situation seems to preclude the possibility of any difference upon which the distinction might rest.

From the perspective of the present paper, this problem is an illusion created by superimposing radical translation onto actual translation. The conceptual distinctions, on which framing rival hypotheses in a case of radical translation depends, derive from the intuitively recognized differences in meaning between "rabbit,"

"rabbit stage," and "undetached rabbit part." The sense of a distinction without a difference comes from the fact that these differences cannot exist in radical translation, so they recede into the background when radical translation is identified with the extreme case of actual translation. Yet the intuitive recognition of them stays with us, and re-emerges in connection with the indeterminacy thesis to give content to talk of "divergent translations" and "incompatible translations" (WO, p. 27). As a consequence, we get the curious duck/rabbit shift. The illusion disappears if we do not identify radical translation with a case of actual translation.

Another problem which makes life with indeterminacy less than idyllic is that translational indeterminacy is a slippery slope. Quine's argument from the absence of "independent controls" in the case of translation between languages, if sound, would also show that translation between dialects of a language must be indeterminate. Further, the same argument would show that translations between idiolects of a dialect must be indeterminate. Nor does the slide stop here. We also have to accept indeterminacy in the case of stages of the same idiolect. Hence, in accepting the initial indeterminacy argument, we are buying a linguistic solipsism of the moment: one's own words of other moments stand to one's words of the present in exactly the relation that "gavagai" is supposed to stand to "rabbit," "rabbit stage," and "undetached rabbit part." Quine himself might be willing to live with indeterminacy so close to home, but few others have been happy with the prospect.31 The solution is simply not to take the fatal first step onto the slope, the step of conceding indeterminacy in translation between languages. The present paper explains why there is no need to take it.32

³¹ See "Ontological Relativity," in *Ontological Relativity and Other Essays* (New York: Columbia 1969), p. 46, where he remarks: "On deeper reflection, radical translation begins at home." The full force of such deeper reflection is brought out by the "theorem" that Hilary Putnam proves from indeterminacy in his *Reason, Truth, and History* (New York: Cambridge, 1981), pp. 22–48. A side benefit of my refutation of indeterminacy is that one no longer needs to accept Putnam's proof

As Quine puts it: "By pressing from below I mean pressing whatever arguments for indeterminacy of translation can be based on inscrutability of terms. I suppose that Harman's example regarding natural numbers comes under this head, theoretical though it is. It is that the sentence '3 ϵ 5' goes into a true sentence of set theory under von Neumann's way of construing natural numbers, but goes into a false one under Zermelo's way" ("On the Reasons for Indeterminacy of Translation," p. 183).

But Quine is wrong in thinking that the nature of the choice between the two ways of construing natural numbers puts pressure on us to recognize indeterminacy. The

Quine is widely regarded as having terminated the Fregean program of taking senses to mediate between word and object. Those of us who still keep the intensionalist faith are frequently made to feel only slightly less benighted than a philosopher who might still be pursuing Hilbert's program. But, if my criticism in this paper is correct, Quine's indeterminacy thesis does not put an end to the intensionalist tradition. Quine cannot be included among those who, like Kurt Gödel, terminated an entire philosophical program. Rather, Quine belongs among the philosophical skeptics who, like Hume, forced subsequent philosophizing to become far clearer about fundamental concepts. Just as Hume's *Treatise* did not eliminate the concept of causality, so Quine's *Word and Object* does not eliminate the concept of meaning. Yet, just as discussions of causality have not been the same since Hume, so discussions of meaning have not been the same since Quine.

JERROLD J. KATZ

Graduate Center, City University of New York

consideration he raises is relevant if, and only if, it has already been shown that the analytic-synthetic distinction cannot be drawn. If the distinction can be drawn, pressing from below is ineffective. For it can hardly matter to questions of translation, which are on the analytic side, that a mathematical sentence comes out true within one set of synthetic statements and false within another. On the other hand, if the distinction cannot be drawn, pressing from below is unnecessary. For then, as we have already seen, the gap in Quine's argument in Word and Object is filled, and the argument goes through without a hitch. Hence, my criticism of Quine's argument in "Two Dogmas of Empiricism" is as well a criticism of his claim that examples like Harman's are a consideration in favor of indeterminacy.

ON FEELING ANGRY AND ELATED

OWEVER it is to be described, feeling angry is not much like feeling elated. Whether these feelings are constituents of emotions or sometimes accompaniments, certain feelings seem relevant to particular emotions, whereas certain other feelings seem irrelevant to the same emotions and, sometimes, irrelevant to any emotion whatsoever.

The appropriate description of these features is one issue. A prior issue, the one to be discussed here, concerns the basis or bases of specification for such noticed differences in the identity of feelings. In the following, I consider possible bases, defend one, and fend off objections to the solution proposed. I shall begin with some assumptions.

A relatively safe assumption is that specifiable cognitive states, including evaluations, are constituents of each and every emotion. For example, whereas anger involves judgments that the object of anger has frustrated one's aims or insulted one in some fashion, elation does not, but rather involves judgments about the exceptional way things have gone, are going, or are expected to go. Since neither the above mentioned judgments nor more intricate complexes of them are thought sufficient for emotions, an assumption only somewhat riskier is that there must be components of emotion additional to these cognitive ones. Whether the additional components are feelings, desires, physiological changes, or some combination thereof remains sufficiently controversial to bar further as-

² See, for example, the criticisms of wholly cognitive analyses by Greenspan, "A Case of Mixed Feelings: Ambivalence and the Logic of Emotion," in Explaining Emotions, Amelie Rorty, ed. (Berkeley: California UP, 1980), pp. 223–250; my "The New View of Emotions," American Philosophical Quarterly, XXII, 5 (1984): 133–141; Lyons, op. cit.; Marks, op. cit.; Jenefer Robinson, "Emotion, Judgment and Desire," this JOURNAL, LXXX, 11 (Nov. 1983): 731–741; and Jerome Shaffer, "An Assessment of Emotion," American Philosophical Quarterly, XX, 2 (1983): 161–172

0022 - 362 X/88/8505/0253 \$ 01.20

This seems to me to be one result—though not an uncontested one—of recent work in the analysis of emotion. See, for example, Errol Bedford, "Emotions," Proceedings of the Aristotelian Society (1956/7): 281–304; William Lyons, Emotion (New York: Cambridge, 1980); Joel Marks, "A Theory of Emotion," Philosophical Studies, XLI (1982): 227–242; Stanley Schachter and Jerome E. Singer, "Cognitive, Social and Physiological Determinants of Emotional States," Psychological Review, LXIX, 5 (1962): 377–399; Robert Solomon, The Passions (Garden City, NJ: Anchor, 1977); but compare William P. Alston, "Emotions and Feelings," Encyclopedia of Philosophy, Paul Edwards, ed. (New York: MacMillan, 1967), pp. 479–486; and Patricia Greenspan, "Emotions as Evaluations," Pacific Philosophical Quarterly, LXII (1984): 158–169.

sumption if we are to consider our question in the least question-

begging fashion possible.

Some general observations about feelings are also in order. Feelings, though always attractive as a constituent of emotion, run into a host of problems which has deterred consideration of them as more than a sometime or frequent accompaniment of emotion.3 Serious discussion of feelings is made difficult because feelings, whatever their precise nature and role, occupy a particularly private part of our psychic life: comparison and contrast is not easy; the lack of subtlety found in the language of feelings is not surprising. Since these difficulties are inherent in the subject matter, we can never enjoy the sort of sophistication in discrimination or comparison concerning feelings which we enjoy with, say, the cognitive components of emotion.

These difficulties, then, bear on the precision attainable concerning the basis for, and specification of, the feelings relevant to emotion, suggesting that it must remain general and in outline. Such generality, however, should not be misconstrued, taken to show that consideration of feelings is misguided, inappropriate, or to be avoided. Rather, it indicates the sort of precision the subject matter admits of, and thereby sets the bounds of reasonable expectations.

Now to our question. Our ability to identify certain feelings, to cite them as relevant or irrelevant to an emotion or emotions, requires that said feelings are specifiable and appropriate. 4 But in what sense and upon what basis?

One possible specification of feelings is in terms of the emotion itself. The feelings relevant to particular emotions are those of that emotion: the feelings relevant to elation are those of elation; the feelings relevant to anger are those of anger, etc.5 Feelings that are not of elation are not relevant to that emotion; feelings that are not of any emotion are feelings which are not relevant to emotion. The basis of specification and relevancy of certain feelings is the emotion the feelings are feelings of. One is able to determine feelings to be of a certain sort on the basis of knowledge of the emotion, and knowledge that said emotion has certain feelings associated with it.

I shall use the 'relevant to' and 'associated with' language in a way which is uncommitted as to whether the identity of the feelings "relevant to" an emotion presupposes or requires that emotion in an important way.

⁵ Hereafter, identification by emotion.

⁵ That feelings are attractive as a constituent of emotion is born out by their frequent role in analyses of emotion prior to this century. Our present century has found such serious problems with that placement that talk of feelings has become something of an embarrassment. But for certain notable exceptions, discussions tend to be dismissive of a place and significance for feelings.

Although as a characterization this may contain some truth, it does not provide an adequate basis to specify the feelings involved in particular emotions, or to exclude feelings as relevant to particular emotions or any emotions whatever. Quite aside from the vagaries of identification by emotion, specifying feelings by emotion is helpful only if feelings are something incidental to and not strictly a part of emotion. Now, this role for feelings may prove true. But, since identification by emotion simply begs the question against certain theories of emotion and feeling (i.e., theories of emotion and feelings in which feelings stand as an independent component), we must look to see whether a more neutral basis for specification is possible. If one is, then a prominent role for feelings in and out of emotion becomes plausible; if one is not, then doubts about any prominent role for feelings must surely be strengthened.

A promising line of inquiry involves sifting through the various senses of 'feeling' distinguishing perceptual, exploratory, propensitory, localized sensation, mock, propositional, etc., senses of the term and isolating a sense or senses relevant to the emotions. 6 William Alston finds a sense, the general condition, which covers four further senses, one of which seems helpful in picking out the sorts of feelings most plausibly involved in the emotions.7 Such delimitation by elimination, though unattractive inasmuch as it does not so much characterize as distinguish the relevant feelings, does help narrow down the feelings relevant to emotion.8

One way to augment the account would be to allow that any feeling from this narrowed grouping is relevant to any emotion, so long as it is conjoined with, associated with, etc., the appropriate cognitions.9 Emotions are, at least, cognitive complexes, ones which may have

⁶ Alston [op. cit.; and "Feelings," Philosophical Review, LXXVIII (1979): 3-34] and Gilbert Ryle ["Feelings," Philosophical Quarterly, I, 3 (1951): 193-205] have developed this line of thought most usefully. Hereafter, I shall refer to this as the sifting method.

According to Alston, under general conditions we have (a) emotional, (b) mood, (c) general bodily conditions, and (d) behavioral tendencies. I suggest that one difference between the feelings associated with behavioral tendencies and those of emotion is the former's prima facie link with behavior and the latter's absence of such a link. A difference between feelings of general bodily conditions and the feelings relevant to emotions is that the former are not intentional, whereas distinctively tively emotional feelings are intentional. Perhaps one can draw the same contrast between mood feelings and emotional feelings. The nonintentionality of mood feelings and emotional feelings. feelings, however, is more controversial. One may have to make room for both emotional and mood feelings as feelings relevant to emotion.

I do not mean to suggest that the last word on this project has been written, only that the project does seem to be identifying a sense of 'feeling' which allows us to distinguish the feelings particularly relevant to the emotions.

Hereafter, conjunct identification.

feelings involved or associated with them; particular emotions are distinct from one another in terms of differences in cognition; particular kinds of emotion have no distinct feelings associated with them other than that they are among the group of feelings deemed relevant to emotion. Any further differences of feeling among emotions would be wholly accidental.

Certainly some will feel that this is as far as one can push the matter. But it has not solved our problem about feelings. For even if we are relatively happy that the feelings relevant to emotion had been or could be separated off in the sifting fashion, still one must specify the feelings that are relevant to particular emotions, say, feeling angry versus feeling elated. Eliminating sorts of feeling as relevant to emotion does not itself provide a basis for differentiating among those feelings which are relevant to emotion. Nor does conjunct identification further the matter. Quite the contrary: conjunct identification has not taken seriously the thought that different feelings associated with certain emotions are recognizably different and appropriate to certain emotions, much less does it provide a basis for this. As we have seen, however, some basis is needed. 10

A better way to augment the sifting method would be to suggest that of the group of feelings deemed relevant to emotion (as determined by the sifting method), the feelings relevant to particular emotions are the ones standardly caused by or typically conjoined with the independently specifiable cognitions constitutive of particular emotions. As a basis this would allow us to say that, since being elated standardly feels in one way and being angry in a quite different way, then, should a feeling like that of elation arise, we can identify it as such—even if it should arise in a context normally associated with a different or no emotion. The feeling remains one of elation. For it is the sort of feeling that is standardly caused by elation.

This explanation is a significant improvement because it does provide a basis for discrimination among the feelings relevant to emotion, including some and excluding others as relevant to particular emotions. It thereby provides a basis for the opening observation that feeling angry is not much like feeling elated.

As it stands, however, there is something of a problem. For speak-

¹¹ This seems to be Anthony Kenny's [Action, Emotion and Will (Boston: Routledge & Kegan Paul, 1963)] position on feeling and emotion. Hereafter, standard cause identification.

¹⁰ In "Feelings and Emotion" [Review of Metaphysics, XXXVII (1984): 303–320], I argue, in effect, both that many hold that the solution of conjunct identification is about as much as one can say about the role of feelings, and that that thesis is too modest. Here I suggest how a stronger relationship might be justified.

ing of feelings "standardly" caused or "typically" conjoined seems to have some statistical, empirical content. What, then, if people should "standardly" come to feel as they do when elated, but in the context of evaluations relevant to anger, and vice versa? Should we allow that, due to possible changes in the regularities of the universe, angry feelings would come to be those previously known as feelings of elation, and vice versa?

Followers of Hume would be prepared to accept this consequence. According to Hume, pride is a pleasant passion, whereas hatred is a painful one. There is no reason for this; it is how we find the world. Presumably, if we were to find the world otherwise, painful feelings could be feelings of pride, while pleasant feelings could be feelings of hatred.

A position with such extreme consequences is not desirable. For it seems simply absurd that the feelings we ascribe to grief are so contingently connected with grief that grief feelings could conceivably come to be those which we presently ascribe to joy. Instead, let us seek an analysis in which there is not this contingency associated with the nature and role of feelings.

I believe that the standard cause solution can be modified so as to allow us to say that feeling and being angry feel in a certain way whatever the empirical regularities. One, then, is not forced to the contingent view about the nature of feelings, suggested by the Humean. To meet this challenge, however, it is necessary to distinguish an identifying and defining project.

Standard cause seems to provide a noncircular way to identify or pick out the feelings relevant to emotion and particular emotions, allowing that certain feelings are recognizable and relevant to emotions and to particular emotions. The identifying project proposals of standard cause do rely on the world being as it is. Should the world turn Heraclitean in the relevant way, yet should we wish to deny the contingency of feelings as just suggested, we may well have to develop different identifying strategies. But in order to deny that contingency of feelings, a related (though distinct) project of defining or delimiting the feelings relevant to particular emotions must be accomplished, and accomplished in a way which does not rely on the contingencies of this world, or at least not in the way the identifying method does. This type of specification, then, is central to an account of feelings which would allow them to be independently specifiable and nonreliant upon the regularities of experience. How, then, is the defining project to be accomplished?

¹² Henry Laycock brought to my attention this consequence of standard cause.

The proposal, mentioned near the outset, that the feelings relevant to emotions are defined by the emotion can be revived here, but, once again, not for those not wishing to foreclose the question of the relative independence of feeling. A different strategy remains necessary.

An assumption at the outset was that judgments constitutive of emotions are specifiable. One can describe, compare, and contrast the judgments involved in, say, anger, versus elation. One thing notable about many of the constitutive judgments is that they value the world in certain ways, e.g., anger involves judgments of insult, whereas elation involves judgments concerning the exceptional way things are going. Furthermore, feelings seem to be valued in and of themselves. For example, painful feelings, aches, etc.—quite apart from their functional utility, etc.—are valued in a certain way, whereas various sorts of pleasant feelings are valued in quite another way. Whether or not the pain of remorse leads to amends, the pain is negatively valued.

Now, if one grants that the feelings relevant to particular emotions should cohere with the evaluations constitutive of emotions if one is to have a coherent emotion complex, then upon that basis one can justify saying things such as "this feeling is not appropriate to this sort of emotion, whereas this one is." The reason for the lack of appropriateness is that the evaluation constitutive of the emotion and the evaluation of the feeling are out of harmony; and the evaluation constitutive of the emotion is emphasized in setting out the emotion, its feelings, and so forth. The reason for appropriateness is a harmony of feeling and the evaluation constitutive of the emotion. ¹⁴

¹³ Stronger positions about certain feelings and evaluation are taken. For example, the pleasure taken in something is essentially the same as the evaluation of that thing. Such a position would tend to support, but is not necessary to, the project undertaken here. See Jan Narveson, *Morality and Utility* (Baltimore: Johns Hopkins, 1967), p. 65.

¹⁴ Hereafter, evaluational specification. This attempt has a certain kinship with attempts, such as Franz Brentano's [Psychology from the Empirical Standpoint, D. B. Terrell, trans. (Boston: Routledge & Kegan Paul, 1971)], to find basic positive and negative emotions from which all others are to be explained. Both projects help us see that more Humean views about feelings, such as the following by Bedford and feeling is inappropriate or unfitting to a situation. But I find this unintelligible. Feelings do not have a character that makes this realistically possible' (Bedford, op. angry if one is not angry' [Solomon, "Emotions and Choice," Review of Metaphysics, xxvi, 1 (1973): 24].

Thus, when one judges in a fashion appropriate to depression but exquisite feelings of pleasure are had, one does not feel depressed, nor is one depressed—whatever people standardly feel in such situations. The valuation of depression involves a grim world, whereas the feeling of pleasure mentioned has a very positive evaluation. So here there is a lack of harmony in combining a positive valuation with something that has a negative one. For depression, both ought to be negative. Similarly, an amused feeling had upon seeing someone's undeserved pain or distress is not part of pity nor appropriate to the emotion. For the evaluation we give of the amused feeling and the evaluation of an undeserved pain are at odds.¹⁵

The contrariness of evaluation need not be this extreme in order to exclude certain feelings as relevant to particular emotions or to certain sorts. For example, it might be that a feeling is too intense in comparison with the valuation of the judgment. If one supposes that the evaluational difference between fear and terror has to do with the extremities of danger, then to judge something somewhat dangerous, yet to feel incredibly distraught and overwhelmed by it, betrays a certain inappropriateness or lack of harmony such that one is puzzled about just what to make of the person's emotional state. It is not clearly a case of terror or fear; it has affinities with each, but the state remains confused. More relevantly, the feelings are not those of a normal, ordinary fear, but those of terror. Feeling that distraught is to feel terror, not to feel fear. ¹⁶

Notice that evaluational specification does not require the presence of the emotions, say, anger, to determine the presence of the feelings, angry feelings. Nor does it require the presence of the evaluations necessary to the emotion. Rather, to be able to determine that a present feeling is or is not an angry feeling presupposes familiarity with anger and its valuation at some level, but not thereby the presence of the anger or the relevant evaluation.

The way evaluational specification provides a basis for the specification of the felt component can be more complex than indicated so far. Consider the pain of love, or delight had in anger when taking

¹⁵ The example is C. D. Broad's ["Emotion and Sentiment," in *Broad's Critical Essays in Moral Philosophy* (London: Allen & Unwin, 1971), pp. 283–301]. Broad speaks of fitting and appropriate emotions, but leaves the notions unanalyzed. Here I offer a part of that analysis.

he has part of that analysis.

An example which frequently causes us much trouble and which parallels the fear/terror contrast in a rather interesting way is the contrast between being fond of or liking someone and the associated feelings, versus being deeply in love with a person and the associated feelings—though here the difference may not be so much that of intensity, but of the depth and durability of the feelings.

vengeance. How are these to be explained? On the surface of it, love seems to involve a rather positive evaluation and thereby should involve corresponding feelings, whereas anger involves a negative evaluation and thereby should require painful feelings. Indeed, these explanations are appropriate as far as they go.

One way to deal with this apparent incongruity of feeling would be to maintain, say, that the pain sometimes associated with love is had at love's frustration or loss, not as part of the love itself. Similarly, the pleasure associated with anger is a pleasure of taking vengeance as a result of anger, not as part of anger itself. Thus, love remains a certain sort of pleasure; anger, a certain sort of pain. The apparent incongruity of feeling is explained away as a feeling which, strictly speaking, is not a part of the emotion—though related to the emotion in specifiable ways.

One difficulty with this response is that, unless one is prepared to maintain that the person having incongruous feelings necessarily is in a mixed feeling state, it suggests that those enjoying vengeance are no longer angry and those pained in love are no longer in love. A claim of a mixed state or the postulation of discontinuance of emotion, however, do not suffice as interpretations of all the relevant experiences. Love, at times, is sheer pain.

A better solution would be to admit that love involves specifiable sorts of evaluations that would require certain pleasant feelings, but add that, when the aims of the evaluation are frustrated or lost, that frustration or loss is itself valued, which, in turn, explains the appropriate place of being pained while in love. 17 Thus, we say that, other things being equal, love is a pleasure, but that, when it is frustrated or lost, because of the meaning of that frustration or loss, frustrated love (and unrequited love) is painful or involves pain. Thus, if someone tries to say that feelings of love are feelings of pain and misery, even sweet misery—even if it is true that this is the plight of most of us most of the time-we can explain why such claims are misleading. Similarly, we say that, other things being equal, anger is distressing and feels so, but that, when the desire for revenge is entertained, set upon, accomplished, it can be most pleasant. Thus, the experience of anger while taking revenge or even contemplating it can end up being rather different from the experience of anger when that revenge is not occurring, anticipated, or possible. Angry feelings, according to this view, can differ. Even so, we can see why it would be

¹⁷ The solution suggested here seems implicit in Aristotle's remarks about anger in his *Rhetoric*, 1378b1-10.

highly misleading to say, without further ado, that said feelings are rather pleasant.

In such states, the complicating factors may take over and, say, the love becomes sheer pain, or it may not take over and leave one with mixed feelings. Exactly what sort of feeling state obtains will have to do with one's situation and one's focus of attention.

Evaluational specification eliminates a good many feelings as relevant to particular emotions, while supporting others as feelings of various emotions, and does so in a way that does not depend on empirical regularities.

To be sure, evaluational specification offers only very general determinations of the feelings relevant to particular emotions. If my earlier thoughts about precision were correct, this is to be expected and is not itself damaging to the proposals. Still, the generality of the account may be worrisome. Let me anticipate and consider some specific objections.

One objection is that the generality of evaluational specification is evidence for the original alternatives of identification by emotion or standard cause. These, after all, were put to the side as less desirable, not proven wrong. If this objection is correct, feelings cannot be an independent constituent of emotion, but must be parasitic upon the emotion and do rely upon empirical regularities. If correct, the only means to determine that feeling angry is not much like feeling elated is to determine the presence of the emotions or the cognitions and judge what the feelings are upon that basis, or to see whether the feeling is the one that is standardly caused by the emotion or cognition.

Yet I do not think this objection works. For the alternatives for identifying feelings (identification by emotion or standard cause) are no more specific. Setting out the feelings relevant to emotion in terms of the emotion itself (identification by emotion) might have the advantage of trivial truth, but it is not very informative about which among a group of feelings is relevant. For either it simply denies the opening observation that feeling angry and feeling elated were not much alike, or it is an appeal to empirical regularities (standard cause), or it is a shorthand for the evaluational method I recommend. If it is the first, it runs contrary to experience. If it is the second, then, when the issue of changes in regularities is considered, it proves much less specific than evaluational specification. If it is a shorthand for the method suggested here, then the disagreement is verbal, not substantive. Hence, the objection to evaluational specification by virtue of a comparative generality in identification of said

feelings is lost. The alternatives are in no better, indeed worse, position.

A second objection from generality is not comparative in nature. The thought behind this objection is that, whatever the attractions of the described method, any procedure that is so general in result must be of little value. For the evaluational method seems unable to exclude feelings that typify anger as feelings of, say, fear. For the disvaluing involved in the cognitive content of each is, on the surface of it, not able to distinguish feelings of anger from feelings of fear. Yet this is absurd: some differentiation is desirable.

The conclusion that the evaluational method, as given, is not able to differentiate feelings of anger versus those of fear seems correct. Moreover, this consequence does seem somewhat absurd.

One can, of course, respond by pinning some hope to future discriminations. After all, no claim has been made that the discriminations discussed are all the relevant evaluational discriminations. Presumably more will be forthcoming; perhaps some which will make the difference concerning feelings of fear and anger. Again, one can re-affirm that the generality with which feelings relevant to emotion can be delimited might not be the disadvantage the objection implies. For exactly which feelings are and are not relevant to which emotions is rather ill-determined; whether to allow certain feelings as relevant to particular emotions is hard to say with precision. Perhaps, then, a means which allows for certain general determinations, but yet which is not terribly specific and precise, reflects not an absurdity for the method of determination, but an appreciation of the rather wide ranging possibilities open to the feelings relevant to emotions.

Although I take it that there is something to these responses, a hope for future discriminations and the recollection of the ill-determinateness of the territory still does not really dispel all our unease about feelings of fear and anger. Fortunately, a further response is available.

We have distinguished the empirical identifying project of standard cause from a nonempirical, evaluational defining project. So far the two have been kept quite separate. Yet the two projects can be combined. That is, we should be able to say, regarding feelings found in anger versus feelings found in fear, that the feelings relevant to each do differ on the basis of certain empirical regularities. Thus, should a feeling like that found in fear arise, even in the context of anger, one is justifiably able to say that it is not a feeling of anger. The basis for saying so is an appeal to standard cause. Pressed on the conceptual point, whether it could be otherwise, we would

have to agree with the Humean that, given some conceivable regularities in the universe, the feelings we speak of in fear could be considered feelings of anger—though at the moment they are not. What would not be possible, due to evaluational specification, is that the feelings which we presently find in anger or fear could be considered feelings of elation, or vice versa—no matter the regularities in the universe.

Thus, we use the evaluational method to provide a basis for certain discriminations in the identity of feelings, and add standard cause to base further discriminations. The refinements based on standard cause are apt to our world of experience, but do not go beyond this; the distinctions based on evaluations go beyond our experience of the world.

The bifurcating here—the notion that a basis for certain discriminations is limited to our world of experience, whereas the basis for other discriminations transcends this—may seem terribly odd. Perhaps we have avoided one absurdity by embracing another. The following thought experiment may help to dispel that suggestion.

Imagine that we were to be structured in a way which means that all painful feelings were like those we now associate with fear. Then, should we have such feelings in the context of judgments of anger, it seems rather plausible to say that we still have anger and feelings of anger. In such circumstances, feelings of fear versus feelings of anger do not differ, much like, say, joy and elation. In contrast, were we to be restructured in a way that meant that all our feelings were like those we now associate with joy, then should we have the feelings of which we are capable in the context of judgments of anger, it seems that we have a different sort of emotion, rather than anger. Certainly we do not have feelings of anger.

Using the evaluational method for certain discriminations, then augmenting it with standard cause for certain additional discriminations in given contexts seems advantageous. Our bases for specifying the feelings relevant to certain emotions becomes flexible enough to allow for certain alignments to be otherwise than we presently find them, but firm enough to prevent any conceivable alignment calling for a change in the identity of feelings.

Hence, the second objection fails to show any disadvantage in the method recommended. What it has brought out, however, is that there is room to be made for the method of standard cause when it augments evaluational discriminations in restricted contexts. Earlier I suggested that standard cause, because of its empirical content, can not be the whole story. We see now, what was suspected then, that it

Digitized by Arya Samaj Foundation Chennai and eGangotri

is a part of the story. It makes discriminations additional to those made by evaluational specification, but discriminations which depend upon the empirical regularities of the world, the constitution of the subjects, etc. To an extent, then, we should agree with the followers of Hume.

An objection with a similar upshot but different in approach notes that the evaluational method rests on the sifting method. Since the sifting method is a contingent methodology, it is possible that the feelings of anger—though still having to satisfy the determinations of evaluational specification—nonetheless could, in different contingencies, turn out rather differently than they presently do. As the above response suggests, however, properly understood this does not seem an unacceptable consequence.

A different objection from contingencies notes that, in the method espoused, one uses evaluations constitutive of the emotion to provide a basis for the specification of the feeling relevant to the emotion. Assumed is a necessary connection between emotion and evaluation. But perhaps that is not so. If the evaluations relevant to certain emotions are only contingently connected with particular emotions, ultimately the connections between feelings of grief and grief must be wholly contingent.

The suggestion here of a contingent connection between emotion and evaluation is peculiar, and contrary to most recent work in emotion. Thus, the objection seems to create little pressure. But even should it be the case that the connection between emotion and evaluation is contingent, that does not show that the feelings are contingent given the evaluations. Thus, even if it is a contingent matter that grief values the world in a particular way and joy in a very different way, it will not follow that the feelings relevant to grief or joy are contingent given these evaluations.

It would seem that we have bases for specifying the feelings relevant to emotion such that we can make sense of the independence of feeling suggested in my opening remarks about feeling angry and elated. In so doing, we make room for the possibility that feelings, at least some of the time, stand as nondependent constituents of emotion.

Queen's University/Kingston

STEPHEN R. LEIGHTON

BOOK REVIEWS

Epistemology and Cognition. ALVIN GOLDMAN. Cambridge: Harvard University Press, 1986. viii, 437 p. \$27.50.

Alvin Goldman has a vision, a vision of epistemology as a multidisciplinary affair orchestrated by philosophers. Although there will still be some normative business to conduct, and some foundational, relatively autonomous questions for the untutored to fiddle with, philosophers who want to grapple with *substantive* questions, who want to practice *epistemics*, the *new* epistemology, will have to know a lot more than they now do about cognitive psychology. This is so because among the substantive questions are questions having to do with the justificational status of beliefs. Since a belief's justification (it is argued) depends on its etiology, on the kind of processes that actually figured in its production (and preservation), philosophers must concern themselves with the processes people use to generate, fix, and preserve belief.

Goldman spends a lot of time, approximately two hundred pages (the second half of the book), illustrating the *modus operandi* of the new field. Since his announced aim is to "restructure" and "redirect" the field of epistemology, one is left to infer that Goldman wants (expects?) converts to follow his example. If Goldman has his way, then, the next generation of epistemologists will proceed in something like the following way. First, they will learn as much as they can (presumably from psychologists and biologists) about the actual mechanisms and processes that humans use in acquiring, preserving, and changing beliefs. Goldman himself gives useful summaries, with many helpful references, of some of the important work that has already been done in such areas as perception, reasoning, imagery, and memory. Potential students of the new epistemology, or just anyone interested in cognitive science, will find these chapters alone to be worth the price of the book.

Fine so far. What next? Well, next one *evaluates* these basic cognitive processes and their associated products. This is the normative part of epistemics, the part at which philosophers will presumably excel.

Evaluate? Evaluate how? According to what standards? For what purpose?

It should come as no surprise to learn that the answers to these questions are supposed to be in the first half of the book. In "Theoretical Foundations," the first eight chapters, Goldman (among

other things) articulates and defends a reliability theory of justification and knowledge. Roughly speaking, knowledge and justification are products of processes that produce a high (enough) ratio of truths. Power and speed, properties which often compete with reliability, are also relevant values in assessing cognitive processes. So. presumably, and once again following Goldman's example, one is to evaluate our cognitive resources (in so far as one can trust psychologists to tell us what they really are) in terms of their reliability, power. and speed.

Why? Why should a philosopher do this? Why, especially, should a philosopher who does not share Goldman's conception of knowledge and justification do this? Does one have to subscribe to a certain epistemological doctrine, reliabilism, in order to practice epistemology in the new way? Not really. Apart from his particular theory of knowledge and justification, Goldman's arguments for the relevance of psychology (or, more generally, cognitive science) to epistemology, and his interest in assessing our cognitive resources, really come down to the obvious point (53, 57) that, if knowing P requires Q, then, to determine whether knowledge that P exists (or is possible), one must know whether Q is true (or possible). If Q is a contingent statement, as it surely must be for some (if not all) necessary conditions for knowledge, then empirical science (construed broadly as the business of finding out which contingent statements are true) is thereby relevant to epistemology—to whether, for example, skepticism is true. Goldman happens to think a necessary condition for knowledge, one of the Qs for knowing (and being justified in believing) P, is a condition relating to the reliability of the processes generating the belief in P. Hence, he thinks the answers to what he calls substantive questions-What can be known? Is skepticism true? Are we ever justified in believing anything?—require us to determine the reliability of cognitive processes. But, if someone disagrees with him about the Q, Goldman will still insist, rightly enough, that, whatever the value of Q, one needs to determine whether Q is true or not to find the answers to substantive epistemological questions.

But surely no one, at least no one who finds skepticism a genuine challenge, is going to argue about this, about the need to determine whether the necessary conditions for knowledge are satisfied (or satisfiable) in order to determine the existence (or possibility) of knowledge. The question has always been: Just what properties of a person are relevant to whether that person knows? What is necessary? Is it reliability or something else? Or something more? Is it something which only scientific specialists can determine? Or is it something more familiar and obvious, something we can all tell (about ourselves and each other) without special scientific expertise?

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Once these foundational questions are answered (What does it take to know?) the answer to substantive questions (Does anyone know?) is a question about whether anyone has what it takes. If what it takes is something only scientific experts can determine (as Goldman seems to think), then we will obviously need the help of scientific experts to determine whether anyone knows anything. If not (as a lot of other people seem to think), then not. But who is going to argue about the point that knowledge requires our having what it takes, and our having what it takes is, broadly speaking, an empirical matter?

At times (e.g., around p. 96) Goldman makes it sound as though he has stronger reasons for insisting on the importance of psychology to epistemology, reasons which are quite independent of his brand of reliabilism. He argues, convincingly I might add, that a system of rules (he calls them J-rules) specifying when someone is justified must describe, not merely the doxastic (or other) states between which transitions are made, but the basic psychological processes by means of which these transitions are carried out. Reaching a true conclusion from premises that logically imply it, for example, is not enough; for there will always remain a question about the method you used to reach that conclusion. Did you know that the premises implied it? If so, did your derivation rely on this knowledge? So a belief cannot be certified as justified merely by knowing where the believer came from in reaching it (specified in a state-transition rule). One also has to know the process used in transit. This being so, cognitive psychology, the field that explores the human repertoire of basic cognitive processes, is necessarily involved in the articulation of the right set of J-rules, those rules which tell us when a belief is justified. Until cognitive psychology tells us what the basic cognitive processes are, no selection can be made of the ones which should be licensed by a right system of rules.

But why should a right system of rules have to select from the processes we actually use? Are we being given to understand that the right set of rules will be (to use Goldman's language) resource-relative: doing the best one can with what one has got? Why should this be so? Perhaps, as Goldman himself suggests (97), it is not humanly possible to be justified. Perhaps no human psychological process is (if this is really what it takes) reliable enough. If this is a possibility, then the right system of J-rules need not, indeed should not, be determined by the processes actually used to fix belief. But, if this is so, what happens to the argument for the special importance of psychology to epistemology? Why, in our quest to determine the right system of J-rules, do we need psychology to identify the processes we actually employ if the "right" processes need not be among them?

This brings us to the very heart of Goldman's theory of knowledge and justification. We are told that the criterion for a right system of J-rules, those rules which specify when a person is justified, is a truth-ratio criterion. Technical issues aside, a truth-ratio criterion is one which requires of any system of rules that satisfy it to yield some ratio of true beliefs to total beliefs. What ratio of true beliefs to total beliefs? Well, that depends. There are, as Goldman puts it, a family of criteria, all of which have some merit, all of which have special attractions. Although he opts for one particular truth-ratio criterion (more of this in a moment) for certain evaluative purposes, Goldman embraces all members of this family. Some of these, the resource-relative ones, even permit truth ratios less than 50% (105). If you have very meager resources, doing the best you can with what you have got will not get you far. Since these resource-relative criteria, conferring (as they do) justification on the products of processes that usually misfire, are obviously not reliabilist criteria, not all of the criteria Goldman endorses are reliabilist. He nonetheless insists that they are all truth-ratio criteria. This is enough to convince him that they form a "close-knit" family and that there is a "core idea of justifiedness" which his theory therefore captures (58/9).

But the criterion Goldman really prefers, the one he will use in "most of what follows" (105), is an absolute, resource-independent criterion. One virtue of this criterion, he tells us, is that it makes skepticism both serious and credible. Since he wants to take skepticism seriously, this is well and good. But resource-independent criteria, remember, are criteria which specify the required truth ratio that a system of J-rules must yield independently of the actual cognitive resources of aspiring cognizers. These are criteria which do not restrict the selection of right processes, those whose products are justified, from actual processes. So, it seems, we are back to wondering what is the special importance of psychology. If you do not have to know how people behave to determine how they should behave, why (other than determining whether they are behaving as they should) must one know how they behave?

With all the fuss made about systems of J-rules and criteria for their selection, it may come as something of a surprise to readers to discover that not only are no J-rules ever given in this book, but the criteria themselves are so underspecified as to be incapable of application. A J-rule system is right if and only if it describes processes that would result in a truth ratio of beliefs which meets some specified high (greater than .50) threshold (106). Before this criterion can be used to select a system of J-rules, someone has to say what this threshold is. Who is this going to be? Another philosopher? A cognitive psychologist? Also, for a belief to count as knowledge, it must be CC-0. In Public Domain. Gurukul Kangri Collection, Haridwa?

produced by a generally reliable process. *How* reliable must a process be for its product to qualify as knowledge? We are not told. In each case, Goldman's silence is apparently the result of his conviction that there is *no* (nonarbitrary) way of fixing these parameters. Our concepts of knowledge and justification are vague on these dimensions and one should not try to impose more precision than common sense contains (51, 106).

Are our ordinary concepts vague on this dimension? If so, why should anyone take the trouble to specify the parameters that Goldman leaves unspecified? Why should anyone trouble to *apply* these criteria if their application requires the specification of a parameter whose value bears no relation to our ordinary concept of knowledge and justification?

Goldman may be right that justifiedness is a many-faceted normative concept for which there is, at best, a family of related criteria (all of them vague) used for different purposes (some of which, as noted above, do not even require a truth-ratio greater than .50). If this is so, I, for one, have trouble seeing much interest, philosophical or otherwise, in "evaluating" cognitive processes (most of which are not yet known) in terms of such a family. Will not the justifiedness of the process depend on the standard used to evaluate? Will not the standard or standards used in evaluation depend on a variety of contextual factors that have to do with the purposes and interests of the agent (not to mention the purposes and interests of the evaluators)? Goldman himself (124/5) speaks of trade-offs between reliability and other epistemic values (e.g., power, speed, goal-responsiveness). He despairs of fusing these multiple standards into a single measure of justifiedness. This, indeed, is why so many of his own "evaluations" in the second half of the book yield little more than judgments that a conjectured process might be good enough. Whether it is good enough depends on which standard (from among the family of acceptable standards) is actually used, on the goals of the cognizer (and, doubtless, the purposes of the evaluator), on the kind of tradeoffs that may have occurred in order to secure increased speed, power, and goal-sensitivity. Given all this slack in the selection and application of standards, it is hard to see why one should make such a fuss about the operational details of the processes to which the standards are to be applied. Surely the processes being assessed are, at a common-sense level, good enough to qualify some of their products as justified according to some acceptable standard (remember, all that is required of an acceptable standard is that it be a truth-ratio standard). Surely the unobstructed vision of normal adults, the memory of people about recent events, and the inferences we all make about, say, another driver's intentions are sometimes reliable enough to qualify the resultant belief as justified according to some truth-ratio standard. Do we really have to consult

cognitive psychology about this?

Although Goldman may be right about justification (it has always struck me that justification was a much more obviously normative concept than knowledge), I, personally, do not see any evidence that our ordinary concept of knowledge is vague or many-splendoured in the way he suggests. At least the people I talk to are quite clear in their judgments that 90% reliability is just not good enough for knowledge. Neither is 99%. People who hold a ticket in a fair lottery in which 100 tickets have been sold, however justified they might be in thinking they are going to lose, do not know they are going to lose if all they have to go on is the probability (99%) of their losing. So what is vague about this? Or does Goldman think none of us (operating with our ordinary vague concept of knowledge) have any strong intuitions about such cases? In retreating behind vagueness at this crucial stage, Goldman is merely denying, without argument, one of the (historically) strongest motivations for skepticism—the idea, surely inspired in part by our ordinary conception of knowledge, that knowledge requires certainty, an evidential posture that excludes (in some suitable sense of possibility) the possibility of error (see, e.g., p. 30).

Throughout this review, I have harped on Goldman's conception of the relation between epistemology and cognitive psychology. I have done so because one of the chief missions of the book is to restructure epistemology by defending a special view about this relation. It is an important theme in the book, one which is, I think, worth taking seriously. But this is only one theme in a very ambitious book. Goldman also has other, more traditional (i.e., foundational) objectives. Besides articulating and defending (against recent criticisms) a sophisticated version of reliabilism—both for knowledge and justification—he also has a sensible and (I think) convincing defense of realism about truth. For those (mindful of recent attacks on folk psychology) who are suspicious of the role of belief in epistemology, there is even a chapter on the problem of content. The book is, in fact, so rich in detail and argument that it will surely inspire debate for many years to come. If the book does not quite succeed in doing what its author wishes—radically changing the way people do epistemology-it will almost certainly have a beneficial impact on the way epistemology is presently being done.

University of Wisconsin/Madison

FRED DRETSKE

Wittgenstein on Meaning. COLIN McGINN. Oxford: Blackwell, 1984. xiv, 202 p. Cloth 19.95.*

Colin McGinn wrote his book after re-reading Wittgenstein's work on rule following and meaning in the light of Saul Kripke's well-known exposition. He set out with the conviction that Kripke had finally clarified Wittgenstein's work and revealed in it "a set of powerful and challenging arguments, disturbing but difficult to dispose of" (vii). He came to the conclusion that Kripke was substantially wrong about Wittgenstein, and that the arguments Kripke develops are weaker than they originally seemed. The book aims to supply four things: a clear statement of Wittgenstein's views on rules and meaning; a critique of Kripke's account as an exposition of Wittgenstein; an evaluation of Wittgenstein's views; an assessment of the argument Kripke finds in Wittgenstein—that there is an insurmountable skeptical paradox about meaning—and of the skeptical solution he offers.

The chief virtue of the book is its accessibility; it sets out a range of ideas about meaning and intentionality clearly and concisely. It treats Wittgenstein as a philosopher rather than an infallible guru, and, while finding much to agree with, is also prepared to attack what he says. McGinn's criticisms sometimes seem oversimple or misplaced, not reaching the heart of Wittgenstein's position. And the book suffers from a lack of context; it would be helpful to have more of an idea of where Wittgenstein's later views on meaning came from. But it remains a stimulating foil to Kripke's reading, and should be particularly useful to those coming to Wittgenstein for the first time.

On McGinn's interpretation, Wittgenstein has three main negative theses about meaning and a positive picture. The negative theses are these: (i) To mean something by a sign is not to be the subject of an inner state or process; (ii) To understand a sign is not to interpret it in a particular way—"there is a way of grasping a rule which is *not* an *interpretation*" (PI 201); (iii) Using a sign in accordance with a rule is not founded on reasons; to follow a rule is, ultimately, to act blindly. Wittgenstein's positive picture is that "to understand a sign is to have mastery of a technique or custom of using it." To say that under-

^{*}I am grateful to Lucy Baines for discussion on the topics of this review.

The main references to Wittgenstein are *Philosophical Investigations* (New York: Blackwell, 1958) (henceforth PI, followed by section number); and *Remarks on the Foundations of Mathematics*, 3rd ed. (New York: Blackwell, 1978) (henceforth RFM, followed by page number). Kripke's exposition is given in *Wittgenstein on Rules and Private Language* (New York: Blackwell, 1982) (henceforth K, followed by page number).

⁰⁰²²⁻³⁶²X/88/8505/0271\$00.50

standing a sign is a technique is to say that it is an ability to do something, a practical skill; this links understanding to use and to behavior. To say that it is a custom emphasizes that following rules is a feature of actual behavior—not something hidden behind behavior. And the idea of a custom is the idea of something that happens regularly; only in a context in which at least some rules are repeatedly applied does it make sense to talk of there being rules at all (McGinn calls this the multiple-application thesis). The idea of a custom involves a certain naturalism about meaning, for the custom of using a sign in a certain way has no other grounding than that that is the way we find it natural to go on once trained: "what underlies (if that is the word) our practices and customs with signs is our human nature in interaction with our training: this is what explains our unreflectively going on as we do" (85). This naturalism about meaning is akin to Hume's naturalism about causation and induction: for Hume, facts about causation are "grounded" in the mind's natural propensity to form expectations given experienced constant conjunctions; for Wittgenstein, facts about meaning are similarly "grounded" in our propensity to find it natural to use words in certain ways once trained.2

McGinn himself professes a general agreement with the orientation of this account of meaning. In particular, he accepts Wittgenstein's critique of anything resembling a seventeenth century theory-of-ideas-type theory of meaning (he makes the interesting claim that Gottlob Frege's view of understanding as grasp of sense is vulnerable to Wittgenstein's negative arguments [98]). And he agrees that "Wittgenstein's fundamental thesis"—that "meaning rests ultimately on the bedrock of our natural propensities" (138) is the only way to keep a genuinely normative picture of meaning without recourse to a magic theory of how grasp of meaning extends to future uses. But, though sympathetic overall, he has considerable disagreements with Wittgenstein. He rejects Wittgenstein's claim that understanding is not a mental state; what Wittgenstein ought to have said is just that understanding should not be assimilated to experiential mental states like pain. And he contends that the state of understanding is the source of correct use and explains the manifestations of understanding. McGinn's position here is plausible, but he underestimates the possibilities of a Wittgensteinian response. Part of his idea is that understanding causally explains use, a view for which he argues along the lines of standard arguments that be-

² This analogy seems to me to be more illuminating, and more accurate, than the comparison of Wittgenstein's "argument against private rule following" to Hume's "argument against private causation" made by Kripke and others.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

liefs and desires causally explain action.³ But there is a lot more to Wittgensteinian objections to such views than those McGinn considers; he seems to miss the depth of the difference between Wittgenstein's whole way of thinking of mind and action and standard causal views in the philosophy of mind.

McGinn's principle disagreement with Wittgenstein is over the multiple-application thesis: "Grasping a rule is a mental state which is potentially multiply manifested, but the potentiality need not, at least as a matter of conceptual necessity, be actualised" (133). This view depends on a sharp distinction between what is required for S to grasp a rule and what is required for us (or S herself)⁴ to say that S has grasped the rule. If you are going to make such a distinction. you must have something to say about what would make it true that S had grasped a particular rule even though neither that nor any other rule had ever been applied; just to repeat that S simply does grasp the rule, and deny the need to say anything more, would surely be to rely on the sort of "queer fact" that McGinn joins Wittgenstein in rejecting. McGinn thinks he can avoid queer facts by stressing that grasp of a rule is ultimately "grounded" in a physically-based propensity, and that S can have that physically-based propensity even though it is never actually manifested. But this says nothing about what it is for S's physically-based propensity to be a disposition to follow a rule, to apply words meaningfully in this or that way, as opposed to a mere brute disposition to exhibit certain physical behavior. This is one case of a general failure in McGinn's book to make enough of the difference between normative regularities and descriptive, causal, empirical regularities—between following a rule and conforming to a rule. Much of what Wittgenstein says about rule following consists in observations about these differences, and the multiple-application thesis is important precisely in making this distinction; an actual custom of applying rules is part of the background needed if a physically-based propensity is to count as a propensity to use words meaningfully, not simply a propensity to make noises.

Turning to Kripke's exposition, McGinn thinks it is seriously wrong as an interpretation of Wittgenstein on three points, and that the arguments Kripke's Wittgenstein offers on these points are unconvincing anyway. First, the idea that the point of Wittgenstein's

See Proceedings on Actions and Events (New York: Oxford, 1980).

See RFM 334/5. "How often must a human being have added, multiplied, divided, before we can say that he has mastered the technique of these kinds of calculation? And by that I don't mean: how often must he have calculated right in order to convince others that he can calculate? No, I mean: in order to prove it to himself."

discussion is to expose a skeptical paradox which shows that there are no facts about meaning. McGinn objects that Wittgenstein does not embrace the skeptical paradox; his statement of the paradox is a reductio of a particular view of meaning—that grasp of the meaning of a sign always consists in an interpretation—and is intended to demolish that view, not to undermine the very idea that there are facts about meaning. And the skeptical paradox is a bad argument against the factuality of meaning, because it depends on an unwarranted reductionist assumption—that we ought to be able to give an account of facts about meaning in terms of something else. Kripke's failure to find any nonintentional facts in which my meaning one thing rather than another can consist shows, not that there is no fact of the matter about what I mean, but that my meaning what I do is a sui generis fact of which no account can be given in nonintentional terms. This seems to me to be substantially right, though again considerably more needs to be said about what it is to mean something by a word or to possess a concept (and considerably more is said by Wittgenstein) than McGinn offers here.

McGinn's second point concerns Kripke's remarks on dispositional analyses of meaning. Wittgenstein, McGinn thinks, gives no authority for Kripke's discussion of the problem posed for such analyses by the fact that our dispositions are finite and include dispositions to make mistakes. And where Wittgenstein does discuss dispositions and counterfactuals, he is inclined to rely on the very sort of analysis that Kripke attacks. For example:

When you said "I already knew at the time [when I gave the order, that he ought to write 1002 after 1000]" that meant something like: "If I had then been asked what number should be written after 1000, I should have replied '1002' " (PI 187).

As an independent argument against dispositional analyses, Kripke's arguments are not telling either, McGinn thinks. The finiteness of our dispositions is a special problem in mathematics, not a general problem for meaning. And our dispositions to make mistakes only seem worrying if one concentrates on what we are actually disposed to do, rather than on what *capacities* we have; I have the *capacity* to add, even if, through stubbornness or carelessness, I do not respond to a particular actual addition question with the sum of the numbers. These comments have a point, but they miss the real thrust of Kripke's critique of dispositional analyses. That thrust is that, even if my dispositions to respond to pairs of numbers with their sums were infinite and immune from mistake, we still could not define what counts as the right answer to an addition problem in terms of those dispositions: my behavioral dispositions are a descriptive matter—if CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

I calculate 'x + y', their sum is the answer I will give; what counts as the right answer is a normative matter—if I am to calculate 'x + y', then I ought to, or must, give their sum. (See K 23 and K 37 for the idea that this, and not the fallibility and finiteness of our dispositions, is the crucial point.) The impossibility of giving an account of normative or intentional facts in terms of descriptive or nonintentional facts is a central theme in Kripke's discussion (he suggests using it also against functionalism (footnote 24, K 35-37)); this is certainly true to Wittgenstein. McGinn is not unaware of the importance of such points—they lie behind his insistence that facts about meaning are sui generis and irreducible: but he does not give Kripke the credit for noticing them; nor, to repeat the criticism, does he have enough to say about Wittgenstein's treatment of them.

Finally, McGinn disagrees with the arguments developed by Kripke about the role of the community. As an interpretation of Wittgenstein, he thinks that any constitutive link between rule following and a community is wrong. Wittgenstein does not mention communities in his positive account of what it is to follow rules. If anything, he thinks that the practice or custom of an individual can itself give the necessary background for rule following; but the issue of community versus individualistic views of rule following "is foreign to his true concerns: it is simply not an issue with which he is wrestling" (82). And McGinn thinks that community views are equally weak as a philosophical theory about meaning. Once we reject reductionist assumptions, there is no need to bring others' responses into the assertion conditions, let alone the truth conditions, of semantic claims. And consideration of examples seems to show that there is no conceptual bar to the idea of a rule follower who is not, and never has been, part of a community of rule followers, and who can be considered a rule follower without reference to the responses of any community.

On both these points it is far from clear that McGinn is right. Wittgenstein obviously rejects the kind of view on which the responses of the community serve directly as the standard against which to judge my own. But that is only the crudest community view (and not one which Kripke espouses—see K 110–112). If we exploit the difference in the role of an empirical proposition and the rule "into which it is hardened" (see RFM 325), we can see the community's use as playing a part in determining what counts as correct use without thinking that '25 × 25 = 625' means that trained people

 $^{^5}$ E.g., PI 226: "Certainly the propositions, 'Human beings believe that twice two is four' and 'Twice two is four' do not mean the same." RFM 332: " 2 52 = 625' cannot be the empirical proposition that people calculate like that, because 2 52 \neq 626' would in that case not be the proposition that people get not this but another

Digitized by Arya Samaj Foundation Chennai and eGangotri

agree in getting '625' as the answer when multiplying 25 by 25. (Indeed, McGinn's own account must use the same distinction to show how it can be that "what we are by nature inclined to do is what it is that constitutes meaning" (85/6), without '1002 follows 1000' meaning that that is what someone finds it natural to say once trained.) And it is not clear that Wittgenstein is unsympathetic to such an essentially communal view of rules. In his published work, as far as I am aware, he carefully avoids commitment to the possibility of noncommunal rule following: typically, he raises the issue but avoids giving a direct answer. For example, "But what about this consensus-doesn't it mean that one human being by himself could not calculate? Well, one human being could at any rate not calculate just once in his life" (RFM 193). And a passage that seems to suggest support for some sort of community view is the following: "Certainly I can give myself a rule and then follow it. But is it not a rule only for this reason, that it is analogous to what is called 'rule' in human dealings?" (RFM 344). It is, at least, not as obvious as McGinn thinks that communities play no essential role for Wittgenstein in the concept of rule following.

Still, whatever Wittgenstein thinks, why should we insist on the conceptual necessity of a community once we abandon the crude community view of rules? Why will the repeated responses of an individual not do everything that the responses of a community could do? One might mention, as pointers, Donald Davidson's⁶ argument that the concept of belief has a place only in the context of a speech community, and John McDowell's⁷ argument that the solution to Wittgenstein's paradox about meaning—the way of grasping a rule which is not an interpretation—requires a community, because it depends on the idea of an immediate awareness of facts about meaning, an awareness available only to fellow members of a community. Much more obviously needs to be said if those arguments are to carry conviction. But my uneasiness about the thought experiments that McGinn takes to show the coherence of the non-

result; and also it could be true if people did not calculate at all." Also see Zettel, 2nd ed. (New York: Blackwell, 1981), sections 428-431. Here Wittgenstein asks: "Does human agreement decide what is red? Is it decided by appeal to the majority? Were we taught to determine colour in that way?" The answer to all these questions is pretty obviously "No."

⁶ E.g., "Thought and Talk," in *Inquiries into Truth and Interpretation* (New York: Oxford, 1984), pp. 155-170, esp. 169-170.

^{7 &}quot;Wittgenstein on Following a Rule," Synthese, LVIII (1984): 325-363, esp. section 11.

community view makes me think that the required extra could indeed be said.

T. W. CHILD

All Souls College, Oxford

NOTES AND NEWS

The editors of *Iyyun*, a philosophical quarterly published at the Hebrew University of Jerusalem, would like to announce that they will be publishing two issues per year in English. These issues will include, alongside original papers, English summaries of the articles published in the Hebrew issues of *Iyyun*. The first English issue is scheduled for January 1989 and will include articles in the philosophy of mathematics by Charles Parsons (Columbia), William Tait (Chicago), Sylvain Cappell (NYU), Warren Goldfarb (Harvard), Shafi Goldwasser (MIT), and Mark Steiner and Itamar Pitowsky (Jerusalem). Manuscripts for future issues may be submitted to E. M. Zemach, S. H. Bergman Center for Philosophical Studies, Hebrew Univ., 91905, Israel.

International Philosophers for the Prevention of Nuclear Omnicide will present symposium sessions on August 21 & 24 at the Eighteenth World Congress of Philosophy, August 20–27 in Brighton, England. The topic of the sessions will be "Philosophy of Peace in the Nuclear Age: What Must We Do to Prevent Nuclear Omnicide?" Papers or abstracts may be submitted to Secretariat, IPPNO, 1426 Merritt Drive, El Cajon, CA 92020.

The editors of *Midwest Studies in Philosophy* would like to announce that, effective vol. 13, it will be published by the Univ. of Notre Dame Press.

Peter Lang Publishing, Inc., is pleased to announce that it will publish a series of scholarly books entitled *New Perspectives in Philosophical Scholarship: Texts and Issues.* The series will be edited by James Duerlinger (Iowa). Books to be published in the series will fall into one of two divisions: either editions and translations of important texts in the history of Eastern or Western philosophy, or works dealing with issues in contemporary or historical Eastern or Western philosophy. Inquiries may be sent to James Duerlinger, Phil. Dept., Univ. of Iowa, Iowa City, IA 52242.

The Warburg Institute announces the separate publication of the index to *Iter Italicum Accedunt alia Itinera*, a finding list of uncatalogued or incompletely catalogued humanistic manuscripts of the Renaissance in Italian and other libraries, compiled by Paul Oskar Kristeller. The index was compiled by Bonnie D. Kent and Ronald Rainey in collaboration with the author.

The Social Philosophy and Policy Center will sponsor a conference on the Foundations of Moral and Political Philosophy to be held September 22–24 at Bowling Green State University. Participants include: Jean Hampton (Pittsburgh), Judith Jarvis Thomson (MIT), Terence Irwin (Cornell), Russell Hardin (Chicago), Holly Smith (Arizona), Eric Mack (Tulane), Peter Railton (Michigan), Allan Gibbard (Michigan), and Stephen Darwall (Michigan). Further information may be obtained by writing to Kory Tilgner, Social Philosophy and Policy Center, Bowling Green State Univ., OH 43403.

The Department of Philosophy of Colorado State University will hold its Second Annual International Conference, co-sponsored by the Endowment for Applied Philosophy, on September 21–25. This year's conference is entitled "The Origin of the Universe: Philosophical and Scientific Perspectives." Invited speakers will include George Gale, John Leslie, Charles Misner, Milton Munitz, Donald Page, John Polkinghorne, David Schramm, Frank Tipler, and Anthony Zee. Further information may be obtained by writing Richard Kitchener, Phil. Dept., Colorado State Univ., Fort Collins, CT 80523.

The Departments of Government and Philosophy of Georgetown University will hold a conference entitled "Liberalism and the Good" on November 3–5. The featured speakers will include Bruce Ackerman (Yale), Brian Barry (London School of Economics), William Connolly (Johns Hopkins), Amy Gutmann (Princeton), John Langan (Georgetown), Martha Nussbaum (Brown), Stephen Salkever (Bryn Mawr), Kenneth Schmitz (Trinity College), and William Sullivan (La Salle). Inquiries may be directed to Gerald Mara, Georgetown Univ. Graduate School, 302 Intercultural Center, 37th & O Streets, N.W., Washington, DC 20057.

Saint Anselm College will sponsor a symposium in observance of its sentennial year, entitled Faith Seeking Understanding: Learning and the Catholic Tradition, on April 20–23, 1989. The conference will consider Catholic education in all its aspects with a special section on Anselm of Canterbury. Those interested in delivering papers or presenting complete sessions are invited to submit a one-page abstract by September 1 to George C. Berthold, P.O. Box 2278, Saint Anselm College, Manchester, NH 03102-9001.

The Center for Critical Thinking and Moral Critique is hosting the Eighth Annual and Sixth International Conference on Critical Thinking and Educational Reform on August 7–10 at Sonoma State University. The theme of this year's conference will be "Infusing Critical Thinking into Subject-matter Instruction: Kindergarten through Graduate School." In addition, intensive preconference workshops will be held on August 5–6. Inquiries may be directed to the Center for Critical Thinking & Moral Critique, Sonoma State Univ., Rohnert Park, CA 94928.

Digitized by Arya Samaj Foundation Chennai and eGangotri

The Department of Philosophy at Rutgers University is pleased to announce that Peter Godfrey-Smith (Sydney University) is the winner of its International Prize in Philosophy for an outstanding undergraduate paper.

NEW BOOKS: ANTHOLOGIES

- ABRAHAM, WILLIAM J., and STEVEN W. HOLTZER, eds.: The Rationality of Religious Belief: Essays in Honour of Basil Mitchill. New York: Oxford, 1987. viii, 269 p. Cloth \$52.00.
- AGASSI, JOSEPH, and IAN CHARLES JARVIE, eds.: Rationality: The Critical View. Boston: Nijhoff, 1987. xi, 477 p. Cloth \$89.50.
- ANNAS, JULIA, ed.: Oxford Studies in Ancient Philosophy, vol. IV: A Festschrift for J. L. Achrill. New York: Oxford, 1987. ix, 243 p. Cloth \$42.50, paper \$22.50.
- BAIRD, ROBERT M., WILLIAM F. COOPER, ELMER H. DUNCAN, and STUART E. ROSENBAUM, eds.: Contemporary Essays on Greek Ideas: The Kilgore Festschrift. Waco, TX: Baylor, 1987. xxii, 306 p. Cloth \$29.95.
- BEHLER, ERNST, ed.: Philosophy of German Idealism. New York: Continuum, 1987. xxii, 284 p. Cloth \$24.50, paper \$10.95.
- BERNAUER, JAMES W., ed.: Amor Mundi: Explorations in the Faith and Thought of Hannah Arendt. Boston: Nijhoff, 1987. x, 220 p. Cloth \$46.50.
- BORN, RAINER, ed.: Artificial Intelligence: The Case against. New York: St. Martin's, 1987. xxxv, 220 p. Cloth \$29.95.
- BRAND, MYLES, and ROBERT M. HARNISH, eds.: The Representation of Knowledge and Belief. Tucson: Arizona UP, 1987. xviii, 368 p. Cloth \$39.95.
- BRODY, BARUCH A., and H. TRISTAM ENGELHARDT, Jr., eds.: Bioethics: Readings and Cases. Englewood Cliffs, NJ: Prentice-Hall, 1987. xxi, 403 p. Cloth \$33.00.
- BUTTIGIEG, JOSEPH A., ed.: Criticism without Boundaries: Directions and Crosscurrents in Postmodern Critical Theory. Notre Dame: University Press, 1987, viii, 251 p. Cloth \$28.95.
- CASCARDI, ANTHONY J., ed.: Literature and the Question of Philosophy.
 Baltimore: Johns Hopkins, 1987. xvii, 333 p. Cloth \$29.50.
- COLEMAN, JULES, and ELLEN FRANKEL PAUL, eds.: Philosphy and Law. New York: Blackwell, 1987. 234 p. Paper \$14.95.
- DAVIS, HOWARD, ed.: Ethics and Defence: Power and Responsibility in the Nuclear Age. New York: Blackwell, 1987. x, 296 p. Paper \$19.95.
- DEUTSCH, KENNETH L., and WALTER SOFFER, eds.: The Crisis of Liberal Democracy: A Straussian Perspective. Albany: SUNY, 1987. ix, 304 p. Cloth \$44.50, paper \$16.95.
- DURPÉ, JOHN, ed.: The Latest on the Best: Essays on Evolution and Optimality. Cambridge: MIT, 1987. xii, 350 p. Cloth \$27.50.

Digitized by Arya Samaj Foundation Chennai and eGangotri

FEAVER, GEORGE, and FREDERICK ROSEN, eds.: Lives, Liberties and the Public Good. New York: St. Martin's, 1987. ix, 272 p. Cloth \$29.95.

FLEW, ANTONY, ed.: Readings in the Philosophical Problems of Parapsychology. Buffalo: Prometheus, 1987. 376 p. Cloth \$24.95, paper \$16.95.

FLOISTAD, GUTTORM, ed.: Contemporary Philosophy: A New Series, vol. V: African Philosophy. Boston: Nijhoff, 1987. ix, 312 p. Cloth \$82.50.

FRENCH, PETER A., THEODORE E. UEHLING, JR., and HOWARD K. WETTSTEIN, eds.: Midwest Studies in Philosophy, vol XI. Minneapolis: Minnesota UP, 1987. 514 p. Cloth \$49.50, paper \$19.95.

GARFIELD, JAY L., ed.: Modularity in Knowledge: Representation and Natural-Language Understanding. Cambridge: MIT, 1987. ix, 427 p. Cloth

\$35.00.

GARVER, NEWTON, and PETER H. HARE, eds.: Naturalism and Rationality. Buffalo: Prometheus, 1987. vi, 289 p. Cloth \$28.95.

GAVISON, RUTH, ed.: Issues in Contemporary Legal Philosophy: The Influence of H. L. A. Hart. New York: Oxford, 1987. vi, 357 p. Cloth \$67.00.

GINSBERG, ROBERT, ed.: The Philosopher as Writer: The Eighteenth Century. Cranbury, NJ: Susquehanna UP, 1987. 245 p. Cloth \$32.50.

GOWANS, CHRISTOPHER W., ed.: Moral Dilemmas. New York: Oxford, 1987. viii, 302 p. Cloth \$29.95, paper \$10.95.

GOWER, BARRY, ed.: Logical Positivism in Perspective. Totowa, NJ: Barnes & Noble, 1987. viii, 129 p. Cloth \$27.50.

GRAESER, ANDREAS, ed.: Mathematics and Metaphysics in Aristotle. Bern: Haupt, 1987. 333 p.

HARNAD, STEVAN, ed.: Categorical Perception: The Groundwork of Cognition. New York: Cambridge, 1987. x, 599 p. Cloth \$59.50.

HILEY, B. J., and F. DAVID PEAT, eds.: Quantum Implications: Essays in Honour of David Bohm. New York: Routledge & Kegan Paul, 1987. vii, 455 p. Cloth \$49.95.

HOROWITZ, IRVING LOUIS, and H. S. THAYER, eds.: Ethics, Science, and Democracy: The Philosophy of Abraham Edel. New Brunswick, NJ: Transaction, 1987. viii, 318 p. Cloth \$49.95.

HOWIE, JOHN, ed.: Ethical Principles and Practices. Carbondale, IL: Southern Illinois UP, 1987. xvii, 171 p. Cloth \$17.50.

IIDA, MASAYO, STEPHEN WECHSLER, and DRAGA ZEC, eds.: Working Papers in Grammatical Theory and Discourse Structure. Chicago: University Press, 1987. xi, 244 p. Cloth \$35.00, paper \$14.95.

KARGON, ROBERT, and PETER ACHINSTEIN, eds.: Kelvin's Baltimore Lectures and Modern Theoretical Physics: Historical and Philosophical Perspectives. Cambridge: MIT, 1987. x, 547 p. Cloth \$40.00.

KENNEDY, ELLEN, and SUSAN MENDUS, eds.: Women in Western Political Philosophy. New York: St. Martin's, 1987. 215 p. Cloth \$32.50, paper \$12.95.

KENNEDY, LEONARD A., ed.: *Thomistic Papers III*. Notre Dame: University Press, 1987. 140 p. Cloth \$19.95, paper \$10.95.

KITTAY, EVA FEDER, and DIANA T. MEYERS, eds.: Women and Moral Theory. Totowa, NJ: Rowman & Littlefield, 1987. ix, 325 p. Cloth \$33.50, paper \$14.95.

- KORS, ALAN CHARLES, and PAUL J. KORSHIN, eds.: Anticipations of the Enlightenment in England, France, and Germany. Philadelphia: Pennsylvania UP, 1987. viii, 290 p. Cloth \$29.95.
- KRETTEK, THOMAS, ed.: Creativity and Common Sense: Essays in Honor of Paul Weiss. Albany: SUNY, 1987. ix, 319 p. Cloth \$44.50, paper \$14.95.
- KRÜGER, LORENZ, LORRAINE J. DASTON, and MICHAEL HEIDELBERGER, eds.: The Probabilistic Revolution, vol. 1: Ideas in History; vol. 11: Ideas in the Sciences. Cambridge: MIT, 1987. vol. 1: xv, 449 p. vol. 11: xvii, 459 p. Cloth \$60.00.
- KRUSCHWITZ, ROBERT B., and ROBERT C. ROBERTS, eds.: The Virtues: Contemporary Essays on Moral Character. Belmont, CA: Wadsworth, 1987. vii, 263 p. Paper, \$17.75.
- LAMB, DAVID, ed.: Hegel and Modern Philosophy. New York: Methuen. 1987. ix, 262 p. Cloth \$59.95.
- LEPORE, ERNEST, ed.: New Directions in Semantics. London: Academic, 1987. xii, 487 p. Cloth \$68.50.
- MacDONALD, GRAHAM, and CRISPIN WRIGHT, eds.: Fact, Science and Morality: Essays on A. J. Ayer's Language, Truth and Logic. New York: Blackwell, 1987. 314 p. Cloth \$45.00.
- McMURRIN, STERLING M., ed.: Liberty, Equality, and Law. Salt Lake City: Utah UP, 1987. x, 205 p. Paper \$8.95.
- MILLER, DAVID, et al., ed.: The Blackwell Encyclopaedia of Political Thought. New York: Blackwell, 1987. xii, 570 p. Cloth \$60.00.
- MORRIS, THOMAS V., ed.: The Concept of God. New York: Oxford, 1987. vi, 276 p. Cloth \$36.00, paper \$13.95.
- MOSER, PAUL K.: A Priori Knowledge. New York: Oxford, 1987. 222 p. Cloth \$36.00, paper \$13.95.
- NEVILLE, ROBERT C., ed.: New Essays in Metaphysics. Albany: SUNY, 1987. xiv, 321 p. Cloth \$39.50, paper \$12.95.
- PAGDEN, ANTHONY, ed.: The Languages of Political Theory in Early-modern Europe. New York: Cambridge, 1987. xi, 360 p. Cloth \$42.50.
- PANGLE, THOMAS L., ed.: The Roots of Political Philosophy: Ten Forgotten Socratic Dialogues. Ithaca, NY: Cornell, 1987. x, 406 p. Cloth \$44.50, paper \$12.95.
- PANTHAM, THOMAS, and KENNETH L. DEUTSCH, eds.: Political Thought in Modern India. Newbury Park, CA: Sage, 1986. 362 p. Cloth \$25.00.
- PAREKH, BHIKHU, and THOMAS PANTHAM, eds.: Political Discourse: Explorations in Indian and Western Political Thought. Newbury Park, CA: Sage, 1987. 314 p. Cloth \$27.50.
- PARKES, GRAHAM, ed.: Heidegger and Asian Thought. Honolulu: Hawaii UP, 1987. ix, 282 p. Cloth \$25.00.
- PENNOCK, J. ROLAND, and JOHN W. CHAPMAN, eds.: Authority Revisited. New York: NYU, 1987. xii, 344 p. Cloth \$42.50.
- PETTIT, PHILIP, and JOHN McDOWELL, eds.: Subject, Thought, and Context. New York: Oxford, 1987. vi, 300 p. Cloth \$45.00, paper \$17.95.
- POJMAN, LOUIS P., ed.: Philosophy of Religion. Belmont, CA: Wadsworth, 1987. xiii, 537 p. Cloth \$33.75.
- PYLYSHYN, ZENON W., ed.: The Robot's Dilemma: The Frame Problem in

Artificial Intelligence. Norwood, NJ: Ablex, 1987. xi, 156 p. Cloth \$29.50.

RASKIN, MARCUS G., and HERBERT J. BERNSTEIN, eds.: New Ways of Knowing. Totowa, NJ: Rowman & Littlefield, 1987. 293 p. Cloth \$34.95, paper \$15.95.

REULAND, ERIC J., and ALICE G. B. TER MEULEN, eds.: The Representation of (In)definiteness. Cambridge: MIT, 1987. 344 p. Cloth \$35.00.

ROSENTHAL, BERNICE GLATZER, ed.: Nietzsche in Russia. Princeton: University Press, 1987. xvi, 424 p. Cloth \$50.00, paper \$19.95.

RYAN, DENNIS P., ed.: Einstein and the Humanities. Westport, CT: Greenwood, 1987. xi, 214 p. Cloth \$32.95.

SALLIS, JOHN, ed.: Deconstruction and Philosophy: The Texts of Jacques Derrida. Chicago: University Press, 1987. xv, 207 p. Cloth \$24.95.

SCHAPER, EVA, ed.: Pleasure, Preference and Value: Studies in Philosophical Aesthetics. New York: Cambridge, 1987. xi, 172 p. Paper \$12.95.

SINGER, MARCUS G., ed.: American Philosophy. New York: Cambridge, 1987. x, 348 p. Paper \$13.95.

SINGER, PETER, ed.: Applied Ethics. New York: Oxford, 1987. vi, 264 p. Paper \$8.95.

SIRIWARDENA, R., ed.: Equality and the Religious Traditions of Asia. New York: St. Martin's, 1987. 173 p. Cloth \$29.95.

SORABJI, RICHARD, ed.: Philoponus and the Rejection of Aristotelian Science. Ithaca, NY: Cornell, 1987. ix, 253 p. Cloth \$42.50.

SRZEDNICKI, JAN T. L., ed.: Stephan Korner: Philosophical Analysis and Reconstruction: Contributions to Philosophy. Boston: Nijhoff, 1987. xi, 171 p. Cloth \$49.50.

STERN, RAPHAEL, ed.: Theories of the Unconscious and Theories of the Self. Hillsdale, NJ: Analytic, 1987. xiii, 282 p. Cloth \$29.95.

STUHR, JOHN J., ed.: Classical American Philosophy: Essential Readings and Interpretive Essays. New York: Oxford, 1987. xiv, 476 p. Cloth \$39.95, paper \$22.00.

vesey, GODFREY, ed.: Philosophers Ancient and Modern. New York: Cambridge, 1987. v, 315 p. Paper \$13.95.

WAITHE, MARY ELLEN, ed.: A History of Women Philosophers, vol 1: 600 B.C.-500 A.D. Boston: Nijhoff, 1987. xxiv, 229 p. Cloth \$67.00.

WEISS, PAUL: Philosophy In Process: vol. x: April 15, 1984-January 18, 1986. Albany: SUNY, 1987. 393 p. Cloth \$49.50, paper \$19.95.

WELTON, DONN, and HUGH J. SILVERMAN, eds.: Critical and Dialectical Phenomenology. Albany: SUNY, 1987. xxiii, 300 p. Cloth \$44.50, paper \$19.95.

WILEY, NORBERT, ed.: The Marx-Weber Debate. Newbury Park, CA: Sage, 1987. 206 p. Cloth \$28.00, paper \$14.00.

wippel, John F., ed.: Studies in Medieval Philosophy: Studies in Philosophy and the History of Philosophy, vol. xvii. Washington: Catholic UP, 1987. 303 p. Cloth \$36.95.

ZECHA, GERHARD, and PAUL WEINGARTNER, eds.: Conscience: An Interdisciplinary View. Boston: Reidel, 1987. xv, 304 p. Cloth \$59.50.

ZIEMBINSKI, ZYGMUNT, ed.: Polish Contributions to the Theory and Philosophy of Lawn America Dom Rockey, kuld grang Collection, Handwar

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 6, JUNE 1988

Carana .

ETHICAL UNIVERSALISM AND PARTICULARISM*

Y aim in this paper is to show that ethical universalism can justify certain kinds of ethical particularism. As a first approximation, ethical universalism may be defined as the doctrine that all persons ought to be treated with equal and impartial positive consideration for their respective goods or interests. This is a first approximation because, among other things, it leaves open the possibility of conflicts between different persons' interests. Without for the present dealing with this, I want to focus on the much more basic conflict between ethical universalism as a whole and ethical particularism. This is a deeper conflict because, according to ethical particularism, one ought to give preferential consideration to the interests of some persons as against others, including not only oneself but also other persons with whom one has special relationships, such as, for example, the members of one's own family or friendship circle or local community or nation or various other restricted social groups. To try to prove that ethical universalism can justify certain kinds of ethical particularism may therefore seem like trying to square the circle, for it involves showing that impartiality toward all persons can justify partiality toward some persons, that equality of consideration can sanction inequality of consideration. Many philosophers have made such attempts, but, not surprisingly, these have had serious shortcomings. I shall refer to some of them below. Before showing how I think the justification can and should be successfully accomplished, certain preliminary questions must be clari-

^{*}This paper was one of my 1987 Bar-Hillel Lectures presented at Tel Aviv University, Israel. For helpful comments I am indebted to David Heyd and Jacob J. Ross.

⁰⁰²²⁻³⁶²X/88/8506/0283\$02.00

First, why is it so important to try to show that ethical universalism can justify ethical particularism? After all, many supporters of ethical particularism have seemed quite assured of its soundness tout court; they have thought it obvious that one does and should give priority to the interests of one's own family, friends, and other close ties without having to justify such preferential treatment through ethical universalism. These supporters of ethical particularism include not only ethical egoists but various intuitionists, pluralists, communitarians, upholders of the "personal point of view," and others. Hence, the justificatory project of this paper may seem pointless.

There are several important reasons for pursuing the justification. Without it, the domain of morality appears to be marked by seemingly irreconcilable contradictions between precepts that require impartial treatment of all persons and directives that permit or even require a privileged status for some persons as against others. On one plausible, widely accepted definition of "morality," either the particularistic concern for the members of one's own family or friendship group emerges as immoral because partial and biased, or else morality is viewed as irrelevant to and even as opaque concerning persons' deepest loyalties.¹

It might be thought that the contradiction can be avoided by adopting a pluralist position whereby some parts of the moral domain—such as those involving the criminal law or basic economic goods—require a universalist principle, and other parts—such as those bearing on familial relations and other personal ties—require a particularist principle. Just as human values and relations in general are of diverse sorts, so too are the moral principles that are respectively relevant to them.² Such a pluralistic differentiation, however, would simply push the whole problem back a step. If the different "human values and relations" are the independent variables determining the relevance and applicability of diverse moral principles, then how do we determine the moral rightness or legitimacy of the values and relations themselves? After all, these may

¹ See Andrew Oldenquist, "Loyalties," this JOURNAL, LXXIX, 4 (April 1982): 173-193.

² For such a pluralist position, see Michael Walzer, Spheres of Justice (New York: Basic Books, 1983), pp. 3–6. For related doctrines which emphasize irreducible plurality and conflicts among moral values, see Isaiah Berlin, Four Essays on Liberty (New York: Oxford, 1969), pp. li, 167 ff.; Thomas Nagel, "The Fragmentation of Value," in his Mortal Questions (New York: Cambridge, 1979), pp. 128–141; Charles Taylor, "The Diversity of Goods," in his Philosophical Papers, II (New York: Cambridge, 1985), pp. 230–247. A convenient anthology of papers on this topic is Moral Dilemmas, Christopher W. Gowans, ed. (New York: Oxford, 1987).

include various forms of domination, obscurantism, and so forth. If it be replied that the human values and relations in question must first pass certain moral tests, then the whole procedure becomes viciously circular. And, if it be further replied that we just know from our actual social practices which values and relations are right, this incurs the relativist deficiency that different persons and groups may uphold divergent and even conflicting values and social relations. Hence, the pluralist defense against the need to search for a justificatory reconciliation between universalism and particularism is not successful.

If it can be shown that, far from being in irreconcilable opposition, ethical universalism and particularism are so related that the former can and does justify the latter, then a certain important coherence is restored to the moral domain, and particularist ideals and relations are seen to have a rationale that is itself moral rather than being a violation of morality. By this rationale, ethical particularism is given a kind of justificatory grounding that is both deeper and more firmly based than the intuitions and appeals to convention or feeling which have hitherto been its main forms of support. And the rationale can also serve to establish which particularisms are morally justified and which not.

A second question concerns the "particularism" to which I refer. Thus far, I have defined it only by its opposition to ethical universalism, as involving inequality of consideration and hence partiality toward some persons or groups as against others. But, on this broad definition, ethical egoism and the "personal point of view" are also forms of ethical particularism. Without denying this, I shall exclude them from the present inquiry because they are not plausible counterpoises to ethical universalism. Ethical egoism has serious internal inconsistencies which have often been remarked.3 As for the "personal point of view," or the preferential status of an individual's own "commitments," which Bernard Williams has especially emphasized, this is excessively vague, at least in its moral bearings. In holding that a person's deepest "commitments" or "projects" should not be subjected to the requirements set by impartial universalist principles, Williams not only fails to distinguish between morally justifiable and morally unjustifiable kinds of commitments but he

See Brian Medlin, "Ultimate Principles and Ethical Egoism," Australasian Journal of Philosophy, xxxv (1957): 44 ff.; and George R. Carlson, "Ethical Egoism Revisited," American Philosophical Quarterly, x (1973): 25 ff.

4 "A Critique of Utilitarianism," in J. J. C. Smart and Williams, Utilitarianism: York: Cambridge, 1973), pp. 115/6; and Moral Luck (New York: Cambridge, 1981), pp. 12 ff.

also seems to deny both the possibility and perhaps even the desirability of a general distinguishing criterion.⁵ His thesis would, for example, justify complete indifference on the part of the dedicated millionaire miser to the sufferings of impoverished multitudes, as well as the deep hostility of a committed racist which would lead him to ignore a drowning child because it happened to be black. Hence, the ethical particularism whose justification is to be found in ethical universalism cannot include indifferently all preferences for deep personal commitments or projects.

The ethical particularism with which I shall be concerned here, then, is confined to preferences for or partiality toward various groups, ranging from one's family and personal friends to larger pluralities of one's community, one's nation, and so forth. Of course, these groups may also vary greatly both in their members and in their moral status. My aim will be to show how and why certain of these group attachments as against others, with their preferential differentiations, can be justified on the basis of ethical universalism. A further ground for confining the particularism to be considered here to groups as against the individual agent himself is that, on one widely accepted definition of "morality," for a judgment to be a moral one it must take favorable account of the interests of at least some persons other than or in addition to the agent or the speaker.⁶ It is this minimally social dimension of morality which is captured in my present restriction of ethical particularism to groups, as against the agent's exclusive concern for his own interests or projects. The conflict between ethical universalism and particularism can thus be viewed as a conflict within morality itself: Among the various moral possibilities of different ethical particularisms, which, if any, are morally right or justified, and can this moral rightness be established on the basis of ethical universalism?

This brings me to my third, and in some ways most important, set of preliminary questions. If the aim is to show how ethical universalism can justify ethical particularism, then what of ethical universalism itself? Does it not also need justification? And may it not be subject to as much morally relevant variability as we saw in ethical particularism?

⁵ This is a central theme of Williams's book, Ethics and the Limits of Philosophy (Cambridge: Harvard, 1985); see pp. 99 ff., 148 ff., 170 ff., 174 ff. I have reviewed this book in Noûs, XXII (March 1988).

⁶ See William K. Frankena, "Recent Conceptions of Morality," in H. N. Castañeda and George Nakhnikian, eds., Morality and the Language of Conduct (Detroit: Wayne State UP, 1963), pp. 1–24; and "The Concept of Morality," this JOURNAL, LXIII, 21 (Nov. 10, 1966): 688–696.

There are at least three answers. First, even if ethical universalism cannot itself be justified, to show that it can justify certain forms of ethical particularism would serve to establish the internal coherence of the whole field of morality. Second, it is sometimes held that ethical universalism, with its emphasis on impartiality and the equal consideration of interests, constitutes the very definition or concept of moral rightness or of "the moral point of view," so that it neither needs nor admits of any independent moral justification. I am doubtful about this attempt to determine what is morally right by verbal or intuitional legislation. But, even if it is accepted, there would still be a need for some other, more general rational justification of morality (in the sense of moral rightness): Why should one be moral or act morally, as against the other courses of conduct that are available for agents? I have elsewhere provided a detailed argument in answer to this question. The argument undertakes to establish that every agent logically must accept that he must take favorable account of the most important interests of other persons as well as of himself, in that he must acknowledge, on pain of self-contradiction, that all actual or prospective agents have certain equal moral rights. For reasons of space, I omit recapitulation of this argument here; but its availability helps to answer the challenge that ethical universalism itself has to be justified.

A third point must be noted in answer to the question presented above about the possible varieties of ethical universalism. It might be thought that ethical universalism does not have the large variability to which I alluded in the case of ethical particularism, because equality and impartiality have a certain conceptual fixity that makes them immune to variation. There are, however, reasons for doubting this. In the present context it will be sufficient to note that there are at least two importantly different kinds of ethical universalism: utilitarianism and the principle of human rights. Each of these provides for the equal and impartial consideration of interests. But utilitarianism is aggregative, whereas the principle of human rights is distributive. Utilitarianism requires that all persons be regarded equally and impartially as loci of values with a view to ascertaining and producing

⁷ Reason and Morality (Chicago: University Press, 1978), chs. 1–3. For briefer versions of the argument, see Gewirth, Human Rights: Essays on Justification and Applications (Chicago: University Press, 1982), pp. 41–67; "The Epistemology of Human Rights," Social Philosophy and Policy, I, 2 (Spring 1984): 1–24; and "The Justification of Morality," Philosophical Studies, LIII (1988).

See, for example, Douglas Rae, Equalities (Cambridge: Harvard, 1981). Discussions of different kinds of equality go back to Aristotle's distinction between arithmetical" (or absolute) and "geometrical" (or proportional) equality; see Nicomachean Ethics, v. 3-4

the greatest possible overall sum of goods. The principle of human rights, on the other hand, requires that all persons be treated equally and impartially with a view to protecting certain important goods of each person severally, as his or her personal due or entitlement, not with a view to maximizing utility overall.

Whereas many discussions of ethical universalism have focused on utilitarianism at least as much as on the "Kantian" principle of human rights, my present analysis will be confined to the latter, for four reasons. First, the human rights framework, unlike utilitarianism, does not admit as morally possible or desirable that the more important interests of some persons should be sacrificed to the less important interests of many other persons when the overall summing of utilities is seen to require such sacrifice. Because of such potential violations of rights, serious flaws mar the attempts to show that utilitarianism, as a version of ethical universalism, can justify varieties of ethical particularism.9 Second, as we shall see, the human rights framework provides criteria for ascertaining degrees of importance of different goods or interests in ways that are not relative to the variable preferences of diverse protagonists. Third, utilitarianism is indeterminate concerning what policies bearing on particularist relations are required by its universalist criterion of maximizing utility. This indeterminacy is epitomized in the disagreement between William Godwin and Henry Sidgwick on such an issue as whether, when a house is on fire, the general utility is better promoted by rescuing one's parent or some stranger whose "importance to the general weal" is greater, when one cannot rescue both. 10

Fourth, there is an important difference between the *kinds* of justification which are provided for ethical particularism by the principle of human rights and by utilitarianism. This difference also applies to various other types of consequentialism besides utilitarianism. When the utilitarian says that overall utility is maximized if one gives particularistic preference to the interests of oneself and one's family, 11 this justification of particularism is extrinsic, not intrinsic. The particularist goods are not valued for their own sakes but only as

⁹ See, for example, David O. Brink, "Utilitarian Morality and the Personal Point of View," this JOURNAL, LXXXIII, 8 (August 1986): 417–438.

¹¹ See William Paley, *The Principles of Moral and Political Philosophy* (Cambridge: Hilliard & Brown, 1830; first published 1785), bk. III, pt. III, ch. 9, p. 166; Sidgwick, *op. cit.*, pp. 432 ff.

¹⁰ See Godwin, Enquiry Concerning Political Justice, Isaac Kramnick, ed. (Harmondsworth: Penguin, 1976), bk. II, ch. 2, pp. 169–171; and Sidgwick, The Methods of Ethics, 7th ed. (London: Macmillan, 1907), pp. 432 ff. The quoted words are from Godwin.

causal, calculated means to the universalist end of maximizing the total amount of utility. A similar point applies to those "Kantians" whose consequentialism is distributive rather than aggregative. Thus, John Rawls's "difference principle" justifies the unequal distribution of economic goods insofar as this is "to everyone's advantage," and Ronald Dworkin holds that "equality of concern and respect" may justify certain "inequalities in goods, opportunities, and liberties." In each of these cases, the particularist preferences or inequalities are justified not as being good or right in themselves, but only as means to the universalist end of advancing some kind of overall equality taken as a fundamental value. Such universalist justifications of particularism thus do not satisfy the contentions of particularists who hold that certain kinds of preferential status—for family, friends, or country—are desirable in themselves, not merely as means to universalist ends.

I shall try to show here that, unlike these consequentialist arguments, the principle of human rights can provide intrinsic justifications of certain kinds of ethical particularism. A further distinction of kinds of justification should also be noted. To justify something is to show or establish that it is right or correct according to some relevant criterion. But the adjective 'right' may signify either what is required or what is permitted; hence, justification too may establish what is right in either of these senses. Now, my project of showing that ethical universalism can justify certain kinds of ethical particularism can be successful even if the kind of moral rightness it establishes for ethical particularism is only that it is permitted by ethical universalism, not required. Nevertheless, the argument to be presented here will uphold, for the most part, the stronger thesis that ethical universalism, in the form of the principle of human rights, can show that certain kinds of ethical particularism are morally required or mandatory, and this as a matter of intrinsic value.

For this purpose, I shall use an elucidation of human rights which I have worked out in considerable detail elsewhere. ¹⁴ I here confine myself to a brief summary. The main point is that human rights are of fundamental importance for morality, because their purpose is to protect equally the general abilities of agency on the part of all persons. The rights serve this function because their objects, what they are rights to, are the generic features and necessary conditions or goods of action and successful action in general, and because the

¹² A Theory of Justice (Cambridge: Harvard, 1971), p. 62.

¹³ Taking Rights Seriously (Cambridge: Harvard, 1977), p. 273.

¹⁴ Reason and Morality, pp. 52 ff.; Human Rights, pp. 3 ff.

rights involve that all persons have equal dignity in that they have equal claims or entitlements to these goods as their personal due.

In their most general aspect, the necessary goods and generic features of action are freedom and well-being. Freedom is the procedural necessary condition of action; it consists in controlling one's behavior by one's unforced choice while having knowledge of relevant circumstances. Well-being as here understood is the substantive necessary condition of action and generally successful action; it consists in having the general abilities and conditions needed for achieving one's purposes. Such well-being falls into a hierarchy of three levels that are progressively less needed for action. Basic well-being consists in having the essential prerequisites of action, such as life, physical integrity, mental equilibrium. Nonsubtractive well-being consists in having the general abilities and conditions needed for maintaining undiminished one's general level of purpose fulfillment and one's capabilities for particular actions. Additive well-being consists in having the general abilities and conditions needed for increasing one's general level of purpose fulfillment and one's capabilities for particular actions. Examples of nonsubtractive well-being are not being lied to, stolen from, or threatened with violence. Examples of additive well-being are self-esteem, education, and opportunities for earning wealth and income. Of course, not all the rights to these kinds of well-being have governments as their respondents or protectors.

This distinction of levels of well-being already shows how differences in degrees of importance of various human goods or interests can be ascertained in a nonrelativist way. And it also indicates a significant criterion for resolving conflicts of human rights, whether between rights of one person or of different persons. For, since the purpose of the human rights is to protect or secure persons' abilities of agency, those rights whose objects are more important because more needed for action take precedence over rights whose objects are less needed for action. This is why, for example, the rights not to be stolen from or lied to are overridden by the rights not to starve or be killed when these respective rights are in conflict. Thus, the rights are prima facie. Whether there are any absolute rights is a separate question. ¹⁵

¹⁵ For a more extensive discussion of how conflicts of rights are to be resolved, see Reason and Morality, pp. 338–354. I have discussed the question of absolute rights in "Are There Any Absolute Rights?," Philosophical Quarterly, XXXI (Jan-Rights, J. Waldron, ed., Oxford Readings in Philosophy (New York: Oxford, 1984), pp. 91–109], and "There Are Absolute Rights," Philosophical Quarterly, XXXII (October 1982): 348–353.

The general principle of human rights is that all persons have equal rights to freedom and to well-being in its three levels. This is obviously an ethical universalist principle. I also call it the *Principle* of Generic Consistency (PGC), because the argument that justifies it combines the material consideration of the generic features or goods of action with the formal consideration of the avoidance of self-contradiction on the part of every agent. Since the rights upheld by the PGC are claim rights that are universally distributed among all persons as actual or prospective agents, the rights entail that all persons also have correlative duties or responsibilities toward all other persons. In the first instance, these duties are negative: each person must refrain from removing or interfering with other persons' freedom and well-being. But the duties are also positive, requiring active assistance in situations where one person can help another to avoid drowning, starvation, or other threats to his basic well-being without comparable cost to himself. In broader social contexts where basic well-being and equality of opportunity (which involves additive rights) can be fostered by collective action, the positive duties require appropriate governmental provisions so far as practicable.

The main thesis of this paper is that the PGC, as the principle of ethical universalism, serves to justify certain kinds of ethical particularism and to show which kinds are morally justified and which are not. But here we run up against the problem I cited at the outset. How can a principle which says that all persons have equal rights justify treating some persons preferentially and hence unequally? The answer might be thought to lie in the fact that rights do not exhaust the justified treatment of persons, so that persons may be treated unequally in those areas of conduct which lie beyond rights, such as where supererogatory conduct is or may be appropriate. This answer, however, would leave it mysterious how the principle of human rights can justify such unequal treatment in the sense of establishing its moral mandatoriness, as against simply leaving it open as a moral possibility.

The answer to our problem is to be found in the fact that the PGC justifies or grounds not only individual duty-fulfilling actions but also certain kinds of social rules and institutions. For the PGC has two kinds of applications: direct and indirect. In the direct applications, the PGC's requirements (of fulfilling the duties that are correlative to the human rights) are imposed directly on the actions of individual persons. Thus, for example, the moral rules that make it a duty to refrain from killing or stealing are direct applications of the PGC, since these duties are direct correlatives of the rights to life and to property. In the indirect applications of the PGC, on the other

hand, its requirements are imposed, in the first instance, on the rules of social institutions, and the requirements of these rules are then, in turn, imposed on the actions of individual persons. According to the indirect applications, social rules and institutions are morally justified if they express or protect persons' equal freedom and wellbeing, in ways to be more fully indicated below.

It is through such indirect applications that the PGC serves to justify various kinds of ethical particularism. The general point of the justification is that the social groups and institutions justified by the PGC involve differentiations of roles and status which require certain preferential differences of treatment. If impartiality of consideration and equality of rights meant that no such differentiations were permitted, then not only would the right to freedom in the formation of voluntary associations be severely limited but such an institution as the criminal law would also be illegitimate. But, on the contrary, the principle of equal rights to freedom and well-being authorizes certain social groupings and institutions wherein persons make differential use of their equal freedom and are subjected to differential treatment because of the ways their actions affect other persons' equal rights to basic well-being. It is because of these authorized differentiations that the ethical universalist principle justifies certain kinds of ethical particularism.

III. HOW THE PRINCIPLE OF HUMAN RIGHTS JUSTIFIES FAMILIAL PREFERENCES

To see more specifically how this justification works, we must look at some of the PGC's indirect applications. Consider, first, the universal right to freedom. This right entails the universal right to form voluntary groups or associations whereby persons freely band together for various purposes, so that all their members voluntarily consent to belong to the respective groups and to obey their rules. Such associations are morally justified, at least in the sense of permitted, insofar as their purposes do not involve violating the human rights of other persons by adversely affecting their general freedom or well-being in ways not derived from the rights themselves. Thus, voluntary associations of murderers and robbers are not justified by the PGC, but baseball teams, colleges and universities, religious groups, and so forth are justified as applications of the universal right to freedom. Because they are justified in the sense of being permitted, they are also justified in the sense that noninterference with them is mandatory.

Two preliminary points must be noted about such voluntary associations or groups. First, although they are applications of a universal human right, they may involve various sorts of particularist ex-

clusion. Membership in the group is voluntary in that all the members freely consent to belong and to conform to the rules, but. by virtue of the purposes for which the group is formed, other persons may be excluded from membership even against their will. This does not, however, prevent them from forming their own voluntary associations for their own purposes, so that in this way the universal right to freedom is still maintained. Moral problems arise here only when the associations in question affect important phases of well-being such that nonmembers suffer serious decreases in their own well-being. Examples of such associations may include certain labor unions, private clubs in which membership is a necessary condition for certain kinds of business advancement, and so forth. Whereas the whole issue is too complex to discuss here in the requisite detail, an important consideration about such strategically important groups is the universalist one that, within the limits set by the conditions for successfully pursuing the economic or other relevant purposes for which the association is formed, there must be no discrimination against persons who fulfill the general criteria for membership as determined by those purposes.

A second point about voluntary associations is that, since they and their rules are indirect applications of the PGC, the PGC's requirements of protecting freedom and well-being are imposed directly on the associations and their rules themselves, not on the actions of individual persons. From this it follows that actions performed in accordance with the rules of such justified voluntary associations may infringe, but not violate, the rights that are upheld by the PGC's direct applications. For example, the umpire in a baseball game may call the batter out and force him to leave the batter's box even against his will. The umpire's action does not violate the batter's right to freedom because the umpire is acting in accordance with rules that are themselves justified by the right to freedom on the part of all the participants, since they have freely consented to play the game and to abide by its rules. It is to the formation of the gameplaying groups and their rules that the PGC's right to freedom is applied here, not to the individual agents or actions within the justified game. Such justified infringement of the batter's right to freedom is different from the violations (i.e., unjustified infringements) of rights committed by robber groups and the like because, among other things, the latter and their rules or modes of conduct do not rest on the voluntary agreement of their victims. In addition, the purposes for which baseball teams are organized serve to justify the particularist preferences of the team's members for their own side's interests in winning a game as against the other side. But such particularist preferences are themselves authorized by the universal right

to form voluntary associations.

We are now in a position to see how the PGC's universalist justification of voluntary associations serves, in turn, to justify the particularist preferences for members of one's own family. It is especially important to consider this subject because familial preferences have often been cited as decisive disproofs of ethical universalism. 16 A marital coupling is a kind of voluntary association or grouping that, like other voluntary associations, is justified by the universal right to freedom. But, unlike baseball teams and other voluntary associations, it is formed for purposes of deeply intimate union and extensive mutual concern and support for the participants, purposes which enhance the partners' general abilities of agency. Such couples are families, and with their children they are simply larger families which are also characterized by the parents' special concern and support for their children. These particularist purposes for which families are formed justify the preferential status whereby their members give priority to one another's interests. Although the children have not themselves voluntarily participated in setting up the family, their special concern for their parents and siblings is appropriately viewed as derivative, both morally and psychologically, from the parents' special concern both for one another and for each of their children and, in this way, for the family as a whole.

Thus, we have here a justification of ethical particularism through ethical universalism. The justification may be summarized in three steps. First, the universalist principle of human rights, in its freedom component, justifies the general moral subprinciple that voluntary associations, as defined above, may be established. Second, this subprinciple justifies the formation of families with their special purposes. Third, these purposes, in turn, justify the particularistic, preferential concern that family members have for one another's interests.

This justification is intrinsic, unlike the extrinsic justifications we noted above in the case of utilitarianism and other consequentialist arguments. When the particularism of familial preferences is justified through the universalist principle of the equal human right to freedom, the fact that the justification proceeds through the special purposes for which couples marry shows that these purposes are

¹⁶ See, for example, Alan H. Goldman, *The Moral Foundations of Professional Ethics* (Totowa, NJ: Rowman & Littlefield, 1980), pp. 4–8; Williams, *Moral Luck*, pp. 17/8; Richard Wasserstrom, "Roles and Morality," in *The Good Lawyer*, David Luban, ed. (Totowa, NJ: Rowman & Allanheld, 1983), p. 26.

regarded as intrinsically valuable, not as valuable only through promoting some universalist end. By the same token, the justification is necessary, not contingent, because the right to form families with their preferential relations is an instantiation of the universal right to freedom, not a contingent means to the effectuation of that right. A similar justification applies to such other kinds of particularism as those whereby one favors one's friends and other close associates.

As was indicated above, this justification of familial particularism does not extend to violations of other persons' moral rights. One form of such violation is when family members use their intrafamilial priorities to engage in criminal activities against outsiders. Another form occurs when a family member who holds some official position, such as judge or teacher, uses it to favor another family member by giving him a lighter sentence or a higher grade than the rules of his position require.

But why is such nepotist behavior wrong? If one gives the universalist reply that the official is violating the impartial rules that govern his position, the question remains why such impartiality should take precedence over the particularist intrafamilial favoritism that we have seen to be justified by the special, particularist purposes for which persons form families.¹⁷

The main answer must recur to the general point that the justification of the family and other particularist groups with their preferential relations cannot extend to violating the moral rights of other persons. When family members hold official positions, whether in voluntary associations or in public institutions, the rules governing the positions prescribe the officials' duties. The officials have accepted those duties and the impartiality they involve as part of the roles they have undertaken in accordance with the respective institutions. The duties and the rules on which they are based entail correlative rights on the part of the persons affected by the official positions. These persons may be affected only directly (as in the cases of the ballplayer in relation to the umpire or the student in relation to the teacher) or as members of the general public (as in the case of the criminal law). But, in all such cases, to act against the impartiality required by the respective rules is to violate the rights of other persons, whether they be ballplayers, students, or members of the

¹⁷ This objection also applies against Robert E. Goodin's attempt to justify familial particularism in *Protecting the Vulnerable* (Chicago: University Press, 1985), pp. 70 ff. If special responsibilities are owed to members of one's family because they are especially "vulnerable to" and "dependent upon" one another, then why their own families?

general public. Such violation sets a limit to the otherwise justified preferential status that family members may accord one another.

IV. SOME OBJECTIONS AND REPLIES

The above argument for the justification of the ethical particularism of the family through ethical universalism may be criticized on various grounds. One objection is that the intrafamilial preferences just described are not really justified by the universal principle of human rights since, precisely because of families' particularistic concern, they may involve ignoring the vital needs and indeed the rights of other persons. 18 There are several answers to this objection. Giving priority to the interests of one's spouse and children need not prevent one from being concerned also with the needs and indeed the economic and other rights of other persons. Priority does not entail exclusivity, save in such dire and exceptional circumstances as where one can rescue only one of two or more drowning persons, and one of them is one's own child. When the needs in question are extensive and recurrent, and especially when they have sociological roots, provision for them requires state rather than individual action, and family members are rightly called upon through taxation and other means to support such action. But none of this is antithetical to the universalist principle's justification of having primary concern for the members of one's own family.

A related objection may take the following form. The demands of universalist morality are categorically obligatory because rationally mandatory. But I seem to be saying that these stringent demands may be overridden whenever someone has sufficiently strong personal likes for friends and other close ties. Such overriding, however, would wreak havoc with the very idea of justice and other impartial moral requirements. The answer to this objection must recur to the argument against nepotism presented above. In addition, whenever objects of distribution (such as academic grades, prison sentences, and so forth) are sufficiently important, their distribution should be determined by general, impartial official or institutional rules which prescribe corresponding rights and duties, and these set limits to justified, particularistically preferential actions.

A further objection emphasizes the opposite side: it holds that my conclusion is not particularistic enough. For precisely because the preferential concern for one's family is justified through a universal principle, the justification may be held to blur or distort the vitally important element of *particularity*. This involves that one views one's spouse and one's children (and for that matter other beloved persons) not as members of some general class or group but rather

¹⁸ See the discussion of "amoral familism" in Edward C. Banfield, *The Moral Basis of a Backward Society* (New York: Free Press, 1958), 1958. CC-0. In Public Domain. Gurukul Kangri Collection, Haliowar

simply as one's own. Thus, Williams has commented that, if a man rescues his wife from some dire peril and has as his "motivating thought" not simply "that it was his wife" but "that it was his wife and in situations of this kind it is permissible to save one's wife," then such a man has "one thought too many." The main point, then, is that, in cases of familial particularism and other such preferential relations, no justification beyond pointing to the particularity is either needed or possible. Indeed, it may also be held that the justification goes rather in the opposite direction, from particularism to universalism. For it is love of one's own family and friends which underlies and generates the love of all mankind, not the reverse.

I have two comments on this contention. First, it seems to confuse logical structure with psychological episode. The crucial point about ethical universalism's justification of ethical particularism concerns not how many "thoughts" one has or just how one actually thinks about some particular situation; it is, rather, a matter of how that situation is logically related to universal justifying principles and, hence, how one ought to think when the full structure of justification is involved. But, second, if one holds that no universal justifying principle is relevant here at all, one may be making either an ontological or an epistemological point. The ontological point is one of extreme nominalism: the facts that she is my wife, that these are my children, are entirely unique, not subsumable under any more general descriptive characterization. But this is implausible. Even if one emphasizes the egocentric particularity of 'my', the fact remains that the possessive relation signified by 'my' is one which distributively characterizes many other relations besides my own; the statement using 'my' is a token of a type that can be truly uttered by many other persons. It is indeed the case that none of these other relations are substitutable for mine so far as concerns my affections and loyalties, and indeed my rights and duties. But this nonsubstitutability also characterizes many other my-relations besides my own. Hence, the emphasis on 'my' is compatible with the recognition that the relation in question is subsumable under, and thus derivable from, a more general description that bears on the justificatory basis of the particularist relation.

The objector may also be making an epistemological point. The awareness that this is my wife and that I ought to rescue her in preference to other persons may be regarded as a kind of immediate intuition that neither needs nor admits derivation from some more general premise. Now, it must be acknowledged that much of our moral thinking operates in this way. But there are at least two reasons

Nozick, Philosophical Explanations (Cambridge: Harvard, 1981), pp. 456/7.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

for not leaving it at that. First, as the history of intuitionism shows, many deliverances accepted as intuitions may be dogmatic contentions which turn out to be false or misguided. Second, even if some intuitions are entirely sound and indeed incorrigible, an increased understanding is obtained by seeing how they are related to and derivable from more general principles, especially when the latter have their own rational warrant. This is akin to the point I made near the beginning of this paper about the moral importance of justifying ethical particularism through ethical universalism. Hence, the justification of familial particularism through the universal principle of human rights provides a clarifying derivation rather than a falsification of the particularist precept.

As for the contention that an upholding of ethical particularism generates support for ethical universalism rather than the reverse, this is a matter of contingent psychological causation rather than of justification. The alleged causal sequence may operate in some cases but not in others. Since, in any case, ethical universalism, as the principle that all humans have certain equal rights, is independently justifiable, it has a rational warrant quite apart from ethical particularism, and it can serve to justify and protect the latter even if, in some cases, it may be psychologically generated by certain kinds of particularism.

V. HOW THE PRINCIPLE OF HUMAN RIGHTS JUSTIFIES COMMUNITARIAN PRIORITIES

Let us now turn to another segment of ethical particularism, whereby one upholds certain communitarian priorities. I here use the general words 'communitarian' and 'community' to refer indifferently to territorially circumscribed sociopolitical groupings ranging from one's city or town to one's state and country, although for the most part I shall be dealing with particularistic priorities for one's country. The words 'nationalism' and 'patriotism' are sometimes used for such priorities of concern. It is important to be especially careful here, however, because some forms of patriotism and nationalism may involve violations of human rights, and hence are unjustified. They include such tendencies as national chauvinism and aggressiveness as well as negative policies concerning the relief of starvation in other parts of the world and the admission of immigrants who flee violations of human rights in their own countries. For this reason I shall not use those words. For this reason, too, racism and sexism are not justifiable forms of particularism, because they involve violations of the human rights of members of excluded groups who are regarded as inherently inferior in dignity or as in other ways less deserving of consideration for their freedom and well-being.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Within the limits set by such violations, there are justified kinds of loyalty to one's own country or other community whereby one has special concern for its flourishing both collectively and distributively. The objects of the concern may range from its national security and welfare policies to its political institutions and economic workings, social and cultural arrangements and traditions, and even aesthetic considerations, as well as the communal relationships fostered by living together in a political society. The forms of concern may range from the mandatory but willing payment of taxes to participation in its political and social life and other modes of special support for one's compatriots, all accompanied by feelings of communal loyalty.

Particularistic priority of concern for such interests of one's country and compatriots can be justified by the universalist principle of equal human rights, but through a different component of the principle from the freedom one that justified familial preferences. There is a crucially important respect in which one's country is not a voluntary association, adherence to whose rules is at the option of its members. The respect in question is that one's country, to be morally legitimate and hence deserving of support, must be at least a "minimal state," which is characterized by impartial enforcement of the criminal law and thus equal protection of the freedom and basic well-being of all the inhabitants. The minimal state, then, secures the equal rights to freedom and basic well-being of all persons within its territory and it enforces the mandatory nonviolation of these rights on the part of all such persons. Thus, the minimal state with its criminal law is justified by a universalist principle. The claim rights of each person to receive the equal protections in question are accompanied by the power rights of state officials to practice their legally authorized protection-securing roles and hence to receive appropriate forms of obedience. It is this nonoptionality of observing the criminal law's protections of basic rights and the minimal state's arrangements for its enforcement which most specifically differentiates the state from voluntary associations.20

Just as we saw in connection with voluntary associations that their justifications as indirect applications of the human right to freedom may, in turn, justify infringements of that right's direct applications by officials who operate in accordance with the association's rules, so similar infringements may be justified for officials of the minimal state. For example, a judge who sentences a criminal to prison infringes his rights to both freedom and well-being. Despite this in-

²⁰ I have discussed the minimal state (and also the democratic state referred to below) more fully in *Reason and Morality*, pp. 290–311. For the distinction between claim rights and power rights, see Wesley N. Hohfeld, *Fundamental Legal Conceptions* (New Haven: Yale, 1964), pp. 41 ff.

fringement, the judge's action does not violate the criminal's rights; on the contrary, it is justified because it is in accordance with the rules of the criminal law, which are themselves justified because they are an application of the human right to basic well-being, which the rules serve to protect equally for all persons. In undergoing such adverse treatment, the criminal is still being treated by common, impartial standards whose purpose is to protect all persons' basic rights equally, and he is treated as a rational person who can understand the rightness of the rules by which he is charged and who has the ability to comply with them. In this way, the ethical universalism of the principle of human rights as applied in the criminal law serves to justify the particularism whereby the criminal is treated unequally through the adverse impact on his own freedom and well-being.

This infringement of the criminal's rights is parallel to the ways in which a would-be criminal attacker may even be killed if this is the only way to prevent him from killing some innocent person. The attacker's rights to freedom and well-being are infringed but not violated, because preventing the violation of a right is not itself the violation of a right.

The territorial circumscription of states and their laws is not antithetical to their being justified by the universal principle that all persons' rights to freedom and basic well-being must be equally and impartially secured.²¹ That the minimal state secures rights only for persons living within its territory is a practical limitation deriving from the fact that the state's functions must operate in relation to persons who are physically present in a specific physical area, although the development of international law may help to mitigate this limitation. But the universal principle also justifies that there be a multiplicity of such states so that all persons' rights are equally secured.

There is a question whether a minimal state will provide the equal protection just referred to unless it is also democratic, in that it protects the civil liberties of each of its members and allows these to be used in the political processes of determining which persons are to have legislative and executive authority, including but not limited to the specification and enforcement of the criminal law. The civil liberties are vital parts of the overall freedom and additive well-being which are necessary conditions of generally successful action to which all persons have equal rights, not only because the lack of these liberties severely affects persons' equal dignity and self-esteem but also because without them persons are drastically limited in coping

²¹ See the discussion of the "particularity requirement" in A. John Simmons, Moral Principles and Political Obligation (Princeton: University Press, 1979), pp. 31 ff. I have discussed this requirement in Hunch Relightantwa 248–250.

with the sociopolitical arrangements that circumscribe their pursuits of their purposes. Thus, the democratic state is also justified by the universal principle of human rights through its components of freedom and additive well-being which bear on the equal protection of the civil liberties. A further aspect of this justification is that membership in a democratic state can be active through persons' use of their civil liberties, as against the predominantly passive membership that the minimal state provides. Such active membership is an important part of the rights of agency whose protection is the basis of human rights, because it helps each person to be a self-controlling, self-developing agent who can relate to other persons on a basis of mutual respect and cooperation, in contrast to being a dependent, passive recipient of the agency of others.

Given these universalist justifications of the minimal and democratic state, the state's protection of basic and other rights serves, in turn, to justify the particularist allegiance of its members to its own flourishing in such spheres as those mentioned above. For ease of exposition, I shall confine myself to the minimal state, but with the recognition that appropriate additions concerning the democratic state can also be made. The justification proceeds through effects produced in protection of that basic well-being which is an essential prerequisite of all action and successful action in general. Because of this causal factor, the justification has an element of contingency; but this is not of the same kind as the contingency we noted above in utilitarian and other consequentialist arguments. The contingency bears not on the universalist maximizing of utility or other universalist consequences, but rather on the particularist consideration of the extent to which, for each individual, the state protects his or her basic well-being.

This justification may be summarized in three steps, parallel to the one presented above for familial preferences. First, the universalist principle of human rights, in its component of basic well-being, justifies the general moral subprinciple that minimal states, each operating within a particular territory, may be established. Second, this subprinciple justifies that the state provides equal protection of the basic well-being of all persons within its particular territory. Third, this protection, in turn, justifies the particularistic, preferential concern that each of the state's members has for its particular interests, in recognition of the protection which he or she receives from the state.

This justification, like the one given above for familial preferences, is intrinsic. The particularistic concern for one's own state's interests is justified, at least in the first instance, not because it serves to maximize utility overall or even because it reflects an equal and CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

impartial protection of human rights, but rather because, for each individual, the state protects his or her personal freedom and basic well-being. These are intrinsic goods for each individual. The justification is also necessary in that the rights to these goods are constitutive parts of human rights, as against being simply causal means toward effectuating some universalist end. It may also be true for some persons that their particularist allegiance to their own country rests further not only on other particularistic grounds of communal membership and friendship but also on the universalist ground of the moral rightness of their country's egalitarian protections of human rights more generally. In the latter case, which is not infrequent, the particularism in question still rests on a universalist ground, although here it is the universalism itself that is upheld as at least part of the relevant end or intrinsic good.

VI. CONCLUSION

What I have tried to show in this paper is that the ethical universalist principle of equal human rights can justify various forms of ethical particularism ranging from familial preferences to special concern for one's own country, and that it can also serve to distinguish between justified and unjustified forms of ethical particularism. The justification works because the social rules and institutions authorized by the ethical universalist principle provide for diverse and special roles and modes of treatment, and these include various kinds of partiality and inequalities of treatment. This particularism can proceed without violations of human rights, and its justified forms cannot include such violations.

An important further consequence of this argument is that the realm of morality has a greater unity than is sometimes attributed to it. This unity, however, is complex, in several ways. As I have tried to show, the principle of human rights is unitary in that it is derived from the necessary conditions of action and successful action in general; but, since these conditions comprise both freedom and well-being, with the latter divided into three levels, the principle itself is diversified, while at the same time the way it is derived provides criteria for resolving conflicts of rights. In addition, and of special importance for the above argument, the principle of human rights has two kinds of applications, direct and indirect, and the latter justifies certain kinds of particularism which modify the universalism of the principle itself, while at the same time they derive from that universalism in the ways sketched above. Thus, the various parts of the principle of human rights have an ordered unity that explains why and how a principle of equal and impartial human rights can justify certain kinds of inequality and partiality.

University 66-Chireletic Domain. Gurukul Kangri Collection, Haridwar GEWIRTH

APPORTIONING CAUSAL RESPONSIBILITY*

S this particle's acceleration due more to gravity or to electricity? Classical physics regards this question as well-conceived. We may answer by examining physical details of the system before us and using our knowledge of pertinent laws. Newton's law of gravity tells us how mass produces a gravitational force; Coulomb's law of electricity shows how charge generates an electrical force. Each of these source laws can be connected with the consequence law "F = ma" to determine which force induces the greater component acceleration.1

In this Newtonian case, two questions seem interchangeable: What contribution did gravity (or electricity) make to the particle's acceleration? What difference did gravity (or electricity) make in the particle's acceleration? These questions are simultaneously addressed by investigating two counterfactual questions: How much acceleration would there have been, if the gravitational force had acted, but the electrical force had been absent? How much acceleration would there have been, if the electrical force had acted, but the gravitational force had been absent? Classical particles obey John Stuart Mill's principle of the composition of causes.2 The result of the two forces is just the sum of what each would have achieved, had it acted alone. We see what each contributed by seeing what difference each made in the magnitude of the effect.

For the accelerating Newtonian particle, apportioning causal responsibility is a local matter—we can assess the contributions of gravity and electricity just by discovering physical facts about the particle and the forces that affect it. In this context, the contribution a cause makes and the difference it makes seem to be one and the same issue. It would be natural to see here an example of some universal principle concerning how science apportions causal responsibility: natural, but unwarranted. For careful examination of how causal responsibility is apportioned in another context casts these questions in an entirely new light.

"Is Jane's height due more to her genes or to her environment?"

^{*} I am grateful to Ellery Eells, Brian Skyrms, Peter Woodruff, and to the editors of this JOURNAL for their suggestions. I also thank the National Science Foundation for financial assistance.

The distinction between source law and consequence law is elaborated and policy to the consequence of Selection: Evolutionary applied to evolutionary theory in my The Nature of Selection: Evolutionary Theory in Philosophical Focus (Cambridge: MIT, 1984), ch. 1.

A System of Logic, Ratiocinative and Inductive (New York: Harper, 1859).

Biologists have been taught to regard this question as meaningless. The proper way to formulate a question concerning nature and nurture, they often say, is at the population level. Junk the question about Jane and replace it with something like the following: "In the population of U.S. adults, how much of the variation in height is explainable by genetic variation, and how much by environmental variation?"

The example of Jane's height shows that the question of how much a cause contributes to the effect and of how much difference it makes in the effect are two questions, not one. The latter is answerable, though not locally, whereas the former is not answerable at all. There is no such thing as the way science apportions causal responsibility; rather, we must see how different sciences understand this problem differently, and why they do so. The particle's acceleration (E) is an effect of gravity (C_1) and electricity (C_2) , just as Jane's height (E) is an effect of her genes (C_1) and environment (C_2) . But this parallelism belies the following differences:

	Newtonian Particle (Gravity/Electricity)	Ontogeny (Nature/Nurture)
How much did C_1 , C_2 contribute to E ?	locally answerable	meaningless
How much difference did C_1 , C_2 make in E ?	locally answerable ³	answerable, but not locally
	Questions are equivalent	Questions are not equivalent

The Newtonian approach to apportioning causal responsibility, I have noted, is based on a theory. So, too, is the biologist's approach to the nature/nurture dispute. Or, more accurately, it is based on a technique. Biologists use a statistical method known as the Analysis of Variance (ANOVA) to say which of an array of causal factors explains more and which explains less of the variation in the effect property found in a population. First and foremost, the method is used to explain the variation in height found in the population that Jane inhabits. After describing how this procedure works, I shall argue that ANOVA can be brought to bear on Jane herself, not just on the population she inhabits. The suggestion will be that ANOVA

³ A proviso will be registered later to defend this entry—namely, that the relevant counterfactuals be nonbacktracking.

can be used to describe which causal factor at work in the singleton case made the largest difference in the effect. So my proposal will connect a populational phenomenon (e.g., variation in height among U.S. adults) with a singleton phenomenon (e.g., Jane's height).

After explaining how the analysis of variance applies to the nature/nurture dispute, I shall try to say why apportioning causal responsibility proceeds so differently, depending on whether one analyzes an organism's ontogeny or a particle's acceleration. Why is a particle's acceleration decomposable in a way that Jane's height is not? This will lead back to Mill's principle of the composition of causes, which is much less central to the difference than might first appear. But, before considering how nature and nurture are dissentangled, we must clarify the concept of locality itself.

I. WHAT IS LOCALITY?

The concept of locality requires more clarification than I shall be able to provide here. So, in the end, I shall have to rely on a somewhat intuitive grasp of the question at issue. Nevertheless, a few remarks may help direct the reader's attention to the right issues.

First, I should note that the problem of locality raised by apportioning causal responsibility is quite different from the one usually discussed in physics. The problem of locality in quantum mechanics has to do with what must be true of two token events if one causes the other. Very roughly, the question is whether a physical signal must proceed from one to the other, passing continuously through a series of intervening space-time points.

The question of causal magnitude is quite different from the problem of whether there can be action at a distance. Suppose, just for the sake of argument, that the physicist's question is correctly answered by some suitable thesis of locality. Even so, this would not answer our question about Jane. Granted, there is a continuous path linking her genes and environment, on the one hand, and her subsequent height, on the other. Locality in the physicist's sense is thereby assured, but the question of apportioning causal responsibility has yet to be addressed. We have yet to see what it means to assess the relative contributions of Jane's environment and genes to her height.

A second problem raised by the question of locality is that it must be formulated in such a way that it is nontrivial. Consider the humdrum fact that causes are rarely in themselves sufficient for their effects. The match's being struck caused it to light, even though it

There are several physical formulations which would have to be sorted out here, as John Earman shows in his "What Is Locality?" in Peter Achinstein, ed., Theoretical Physics and Phys cal Physics in the Hundred Years Since Kelvin's Baltimore Lectures (Cambridge:

would not have ignited if the match had been wet or if there had been no oxygen in the air. This is the trivial sense in which locality fails; whether one event will cause another depends on features of the world external to them both.

In the nature/nurture controversy, we acknowledge this familiar point when we say that genes are not in themselves enough. For Jane to reach a certain height, she must be raised in an appropriate environment. Genes are no good, unless supplemented by numerous meals. Nor is environment in itself sufficient, since there are genetic configurations that will impede Jane's growth, no matter how much milk she drinks. The cliché is that development is the result of a gene/environment interaction.

Be this as it may, our problem is not thereby resolved. Neither genes nor environment are themselves sufficient. But, when they conspire to produce Jane's height, how are we to assess their relative contributions? In particular, is this matter resolvable by attending to features of Jane's physical development from zygote to adult? Or must we look to features outside of Jane's ontogeny and the genes

and environment that produced it?

It seems clear that the problem of locality depends on how one carves up nature into discrete physical systems. If Jane's sequence of environments and genes comprise the system in question, then locality will be refuted if we find that factors external to them play a role in determining which mattered more. If from the first we think of the population in which Jane resides as the relevant unit of inquiry, however, then we may find that locality is verified. The causal facts about Jane may turn out to depend on the population, but the facts about the population may not depend on anything external to it. If so, whether locality is true or false will turn on whether we pose our question about Jane or about her population. As emphasized earlier, the principal question of interest here concerns "singleton" physical systems (Jane, the particle), not populations of such.

The question of locality presupposes that descriptors of an individual can be divided into ones that are "intrinsic" and ones that are not. Although the predicate 'x lives in a genetically homogeneous population' is true of individuals, it nonetheless expresses a nonlocal property of them. I do not believe that the distinction needed here can be drawn syntactically; nor should it be, since the conclusions reached should be language independent. Nor do I have an informative semantic criterion to suggest. As in other problem areas (e.g., how the thesis of determinism should be formulated), we must rely on a somewhat intuitive grasp of what it is for a property to

be local.

II. THE ANALYSIS OF VARIANCE

Why do biologists think the nature/nurture question is meaningless unless formulated in terms of population variation? To be sure, it is silly to think of Jane's six-foot stature as decomposing into two feet due to genes and four feet due to environment, as if the genes built Jane from the navel up, while the environment took care of what lies below. But one silly suggestion does not show that the question of causal magnitude must be understood nonlocally.

To see why biologists think that apportioning causal responsibility requires a populational analysis, we need to consider the basics of the ANOVA technique. So as to avoid the macabre prospect of experimentally manipulating human beings, let us switch from Jane to a corn plant. It has a certain height; we want to know how genes and environment affected this outcome. As its name suggests, the analysis of variance understands this problem at the population level. Instead of focusing exclusively on the singleton case, we consider different possible corn genotypes (G_1, G_2, \ldots, G_n) and different possible environmental conditions (E_1, E_2, \ldots, E_m) —amount of fertilizer, for example. There are then nm possible "treatments." We might divide a field into different plots, the corn plants on each plot receiving one of the treatments $(G_i \& E_j, i = 1, 2, ..., n \text{ and } j = 1, 2,$..., m). Each plot of land would contain the same number of plants; within a plot, the plants would have identical genotypes and would receive identical amounts of fertilizer. We then would record the average height within each plot and enter the result into an n by m ANOVA table:

		Environmental Variable				
		E_1	E_2	• • •	E_m	Averages
	G_1	x ₁₁	x ₁₂		x_{1m}	$M_{1.}$
Genetic Variable	G_2	x ₂₁	x ₂₂		x_{2m}	M _{2.}
						No. of the last of
	G_n	x_{n1}	x_{n2}	• • •	x_{nm}	$M_{n_{\star}}$
Av	erages	$M_{.1}$	M.2		M _{.m}	

The marginal averages simply record the average height of plants with the same genotype which receive different amounts of fertilizer (the various M_i s) and the average height of plants receiving the same amount of fertilizer, but having different genotypes (the various

 M_{j} s). One final number, not shown in the above table, is the grand mean M, which is simply the average height across all treatments.

It is a matter of arithmetic that the difference between any cell in the above table and the grand mean $(x_{ij} - M)$ must equal the sum of three terms: the so-called *environmental main effect* $(x_{ij} - M_i)$, the so-called *genetic main effect* $(x_{ij} - M_j)$, and the so-called *gene-environment interaction I* (this last term is a fudge factor, which merely insures that the sum comes out right):

$$(x_{ij}-M)=(x_{ij}-M_{.j})+(x_{ij}-M_{i.})+I.$$

Rearranging the above equation a little, we can show how the average height within any treatment plot must be related to the averages defined on the entire field of plots:

$$x_{ij} = M + (x_{ij} - M_{.j}) + (x_{ij} - M_{i.}) + I.$$

What I have set forth so far is quite unobjectionable. But what use is this partitioning when it comes to the task of causal explanation? I begin with a negative claim: ANOVA does *not* explain the occurrence of singleton effects (i.e., why a given plant has a height of x_{ij}). In the next section, something positive will be defended concerning the relevance of ANOVA to explaining singleton phenomena.

It seems quite clear that the height attained within a given plot does not causally depend on the heights attained in other plots. This is true even though the height within a plot is the sum of terms that represent that plot's deviations from various means (and a corrective factor, *I*). I would supplement this causal claim with an explanatory one: If you want to know why the corn plants in a given treatment attained the height they did, do not cite the information codified in the analysis of variance formula. The analysis of variance does not identify causes or explain the upshots found in individual cells.

Consider the fact that a given treatment $(G_i \& E_j)$ can be embedded in different experimental designs. For example, I might investigate plants that have the same genotype and place them in different environments. In this case, I shall find that the genetic main effect is zero. As biologists say, all the variance in height is explained by

⁵ For the statistical details of this method, see Robert Sokal and F. James Rohlf, Biometry: The Principles and Practice of Statistics in Biological Research (San Francisco: Freeman, 1969). I have omitted mention of an "error" term; this plays a role in ANOVA inference, but does not affect the points about causality to be made in what follows.

⁶ The following arguments are explored in somewhat more detail by Richard Lewontin in "The Analysis of Variance and the Analysis of Causes," American Journal of Human Genetics, XXVI (1974): 400–411; and by myself in The Nature of Selection, op. cit.

environmental variance. Likewise, I might see how plants of different genotype fare in a single environment. In this case, the environmental main effect will be zero, and all the variance in height will be explained by genetic variance. This shows how misguided it is to interpret the analysis of variance as explaining why a plant in a given treatment cell attained the height it did. If I run the experiment one way, I shall say that the genes played no role; if I run it in the other way, I shall conclude that the environment played no role. And, if I vary both treatments, as I did initially, I shall doubtless conclude that genes and environment both played a role.

An additional reason for rejecting the analysis of variance as a device for identifying the causes that contributed to a singleton effect is provided by considering what this procedure says about characteristics that are universal in a population. Let us switch from corn-plant height to the human characteristic of having a single head. If we consider a population in which everyone has just one head, the analysis of variance will tell us that the genetic and environmental main effects are zero. It would be a mistake to conclude from this, however, that Jane's environment and genes played no role in providing her with a head.

The calculations derived from an ANOVA table do not allow one to deduce what the causes of a given singleton effect are. But this is not to deny that ANOVA can be used to describe the relative contributions that different factors make to explaining the pattern of variation of a trait in a population. Although Jane's genes obviously contribute to her having a head, it may or may not be true that genetic variation helps explain the way that phenotype varies in the population. This is perhaps why biologists have thought of apportioning causal responsibility as fundamentally a population level problem, not one that can be meaningfully addressed for the singleton case. I now want to suggest, however, that the proper use of ANOVA at the population level has relevance to a question about singleton events. One can say which causal factor at work in a singleton event made the greatest difference by embedding that singleton effect in a population, and then judiciously applying the ANOVA to that population.

Is suggest that causal responsibility for Jane's height can be apportioned between her genes and environment by considering two counterfactuals. How tall would Jane be if she had different genes, but the same environment? How tall would Jane be if she had a different environment, but the same genes? Environment is a more powerful influence than genes, if changing her environment would lead Jane's height to depart more from its actual value than would changing her genes.

To answer these two counterfactuals questions, we must elaborate. If Jane were raised in an environment different from the actual one, what would her environment be like? And, if Jane had a different complement of genes, what genes would she in fact have? My suggestion is that the way we fill in the details here is often nonlocal. When this is so, the question of causal magnitude is also a nonlocal one.

When we consider what genes Jane would have had if she had been differently endowed, we do not consider gazelle genes or Martian genes. In some sense, we go to a more "similar possible world." We may consider the other genes that were available from Jane's parents. If these are the same as the ones she actually possesses (i.e., Jane and her parents are all homozygous for the genes in question). we may go back to Jane's grandparents. Or perhaps, rather than tracing the lineage backwards in time, we may look at the other alleles present in the local population that Jane inhabits.8 If there are only two alleles in this local population and Jane has one of them, then it is natural to say that she would have had the other one, if she had had a gene different from the one she actually possesses. If there are several, then we might calculate Jane's expected height as a weighted average of the heights she would have had with each alternative gene, the weighting being supplied by the genes' population frequencies.

Similarly, when we ask what environment Jane would have inhabited had she not inhabited the one she in fact did, we do not imagine her floating weightlessly in outer space, nourished by a diet of candy bars. This is not a "similar possible world." Since human beings inhabit such an incredible variety of environments, it often will be impossible to say which environment Jane would have inhabited, had she pursued a different way of life. If so, a natural strategy is to do what we did above for the case of multiple alleles. We assign different possible environments a probability and then calculate Jane's expected height as a weighting over these.⁹

⁸ From this point on, I shall imagine that Jane is a haploid organism, so that I can talk of her having a given gene (not two genes) at a locus. This is simply to simplify

⁹ The probabilistic formulation of the counterfactual test advocated here for apportioning causal responsibility is basically the same as the one I describe in "Causal Factors, Causal Inference, Causal Explanation" [Proceedings of the Aristotelian Society, Series B (1986): 97–113] for inferring which of the causal factors present before an effect was its token cause.

⁷ Like all counterfactuals, the two at issue here at times may be vague or indeterminate. To the degree that this is so, I suggest that the question of causal magnitude also is vague or indeterminate.

It seems intuitive that, in both these problems, we elaborate our counterfactuals by attending to features of the world which are extrinsic to Jane's actual genes and environment and the height she actually attained. It is no intrinsic feature of Jane which dictates what genes she would have had, if she had had a different genome. The same holds true of her environment.

We can illustrate this idea in a thought experiment akin to Hilary Putnam's story about twin earth. ¹⁰ Jane lives in environment E_2 and has genotype G_2 . We need to consider what her environment would be, if it differed from E_2 , and what her genotype would be, if it differed from G_2 .

We shall consider two possible contexts. In the first (I), Jane lives in a population in which the only alternative to E_2 is E_1 , and the only alternative to G_2 is G_1 . In the second (II), Jane inhabits a population in which the only alternative to E_2 is E_3 , and the only alternative to G_2 is G_3 . We shall see how this contextual difference leads to contradictory assessments of how we should apportion responsibility for Jane's height between her genes and environment.

The following table describes the height that Jane would have in each of the nine possible gene/environment combinations:

			Environments		
			E_1	E_2	E_3
Genes	G_1	I	x ₁₁	x ₁₂	x ₁₃
	G_2		x ₂₁	x ₂₂	x ₂₃
	G_3		x ₃₁	x ₃₂	x ₃₃

Jane's actual height is given by x_{22} . If she lives in a population of the first type (I), then we assess the relative impact of her genes and environment by comparing x_{12} and x_{21} . If she lives in a population of the second type (II), we must compare x_{23} and x_{32} . Suppose that $x_{22} = x_{12} = x_{23}$. This means that, in situation I, her genes make no difference to her height; if she had had a different genome, she would have had the same height. In situation II, however, it is her environment that makes no difference; if she had had a different

¹⁰ See "The Meaning of 'Meaning'" in Mind, Language and Reality: Philosophical Papers, vol. 2 (New York: Cambridge, 1975), pp. 215-271.

environment, she would have had the same height. So, whether genes matter more than environment depends on features of the world external to Jane's own environment and genes. It follows that the question of causal magnitude is nonlocal.

The above table is just what one would obtain by pursuing an analysis of variance on causal factors antecedently identified. One would clone Jane and place her in different environments and place genetically different individuals in the environment she actually inhabits. Or, more precisely, since ethics prohibits this manipulation, one would try to assess what the consequences would be if these things were done. When we make counterfactuals actual, so to speak, we use the analysis of variance to apportion causal responsibility. 11 IV. COMMENSURABILITY AND LOCALITY

In arguing that the analysis of variance correctly apportions causal responsibility, I mean to suggest that it can identify the difference that various causes make in the observed effect. This is quite different from assessing how much each contributed. This latter question I interpret as a local one, which the facts of the matter about nature and nurture render unanswerable. Why this is so we can see by inventing a science fiction scenario in which the second question is quite intelligible.

Suppose height were the result of the accumulation of height particles, which organisms could obtain from their environment and also from their genes. Imagine that an individual's height is some increasing function of the number of height particles obtained from all sources. If so, we could look at local facts about Jane and say whether her genes or environment contributed more.

Of course, there are no such things as height particles, but this thought experiment suggests the following (somewhat vague) conjecture: For it to make sense to ask what (or how much) a cause contributes to an effect, the various causes must be commensurable in the way they produce their effects. Height particles would provide this common currency.

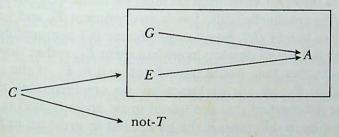
The idea of height particles shows the inadequacy of one diagnosis of why locality fails in the nature/nurture case. In the Newtonian example, we can ask what would happen if no gravitational force acted or if no electrical force were at work. In the nature/nurture

¹¹ When there is a univocal answer to the question of how tall Jane would be if her genes were different, probabilities need not intrude; when there are multiple-alternative possible genomes, a probabilistic weighting is needed. In both these cases, however, there may be disagreement over whether there is a uniquely correct answer to the question of causal magnitude, or whether the "choice of a reference class" is inherently arbitrary. My argument is neutral on this, since, in either case, the thesis of locality fails.

case, no significant answer is obtained by asking how tall Jane would have been if she had had no genes or no environment. The intelligibility of these questions is not essential, however. If there were height particles, the question of whether genes or environment contributed more would make sense, even though an organism requires genes and an environment of some sort if it is to exist at all.

Nevertheless, this difference between gravity/electricity and nature/nurture is not without its significance. I have claimed that the difference made by gravity and by electricity is a local matter; intrinsic features of the physical system under study determine what would happen if either force had acted alone. But we may embellish the details of this example to show that the issue of locality is, even here, not so straightforward.

We so far have considered a physical system made of a particle and the gravitational (G) and electrical (E) forces that affect its acceleration (A). Suppose we embed this physical system into a larger context. Let us imagine that some cause C—a switch, say—insures that a third force T of given magnitude would have come into play if and only if there had been no electrical force (E). The set-up is as follows, with a box drawn around the factors that comprise what I have so far called "the physical system":



There now is an ambiguity in the question of what would have happened if gravity had acted, but electricity had not. We could interpret this, as we have so far, as meaning that gravity would have acted *alone*. Or we could reason that, if electricity had been absent, then the third force (T) would have acted in its stead.

That is, we must decide whether or not we allow counterfactuals to "backtrack." I believe that both interpretations have their place and that it is contextual clues that settle which reading is appropriate. If so, there is no univocal answer to the question of whether the difference made by gravity and electricity is locally determined. If

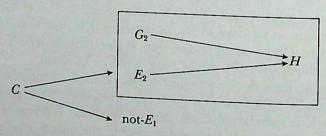
¹² I here use the vocabulary, if not the full-blown theory, developed by David Lewis in his "Counterfactual Dependence and Time's Arrow," *Noûs*, XIII (1979):

Digitized by Arya Samaj Foundation Chennai and eGangotri

we forbid backtracking, then the question of what difference gravity and electricity made reduces to the question of what each would have achieved if it had acted alone. This is a local matter. If we permit backtracking, however, then the question becomes nonlocal. We must consult physical facts extrinsic to the particle and the gravitational and electrical forces acting on it. Thus, the lower-left entry in the table presented in the first section requires that counterfactuals not backtrack.

No similar ambiguity can arise in the nature/nurture case. The reason is that it makes no sense to imagine an individual developing without any environment at all and that an individual will not develop at all, if it has no genes. The ambiguity in the gravity/electricity case was made possible by the fact that we can ask what would have happened if only one of those forces had been present. This provides the nonbacktracking and local reading of the question of what would have happened if one had been present and the other absent.

This local interpretation is not available in the nature/nurture case; we are forced to consider the wider context which dictates what alternative environmental and genetic factors would have been present, if the actual ones had been absent. Here the question of how tall Jane would have been, if she had had a different complement of genes (or a different environment) must be understood in terms of backtracking counterfactuals. Jane's environment E_2 and genes G_2 produce her height (H). Some causal factor (C) presumably determined that she would develop in environment E_2 , rather than in E_1 . If we reason that Jane would have developed in E_1 , if she had not developed in E_2 , we are backtracking; we are imagining that the cause that determined that she would grow up in E_2 and not in E_1 would have been different, if she had failed to grow up in E_2 . The causal facts are shown below:



I conclude that the question of what difference two causal factors made may be local or nonlocal in the Newtonian case, but that it must be nonlocal when the issue concerns nature/nurture.

Thus, there are at least two differences between the Newtonian particle and the real world of nature and nurture. Causal contribu-

tion is well-defined and local in the first, but not in the second. In addition, the two cases part ways over how the difference a cause makes is determined. This can be understood either locally or non-locally in the Newtonian case, but it must be understood nonlocally in the case of nature and nurture.

It may be thought that the key to the difference between the Newtonian and the biological examples lies in the fact that the former obeys Mill's principle of the composition of causes. As mentioned before, source laws describe which physical properties of the particle system generate this or that force, and the consequence law "F = ma" allows the separate contributions of gravity and electricity to be described in a common currency. This generic concept of force is indispensible, if the contributions of each cause are to be comparable. But it now is time to see that it is entirely irrelevant whether the system under study obeys the principle to which Mill called our attention.

Biologists would use the terms 'additive' and 'nonadditive' to draw Mill's distinction between systems which obey the principle of the composition of causes and ones which are, in his terminology, 'heteropathic' (op. cit., p. 213). We now may elaborate our story about height particles to show why additivity is neither necessary nor sufficient for locality.

As long as such particles exist, the relative contribution of genes and environment to height will be locally assessible, regardless of the formula that translates number of particles into inches and feet. Suppose that height increases as the number of particles goes from 0 to 50 and declines thereafter. If 40 particles were obtained from the environment and 40 from the genes, we would assign an equal contribution to each source. And this would be correct even though an individual with 80 particles is not twice as tall as an individual with 40.

This shows why additivity is not necessary for locality. To see that it is not sufficient, we must realize that the analysis of variance has no problem describing an additive relation between genes and environment; the data displayed in an ANOVA table are additive, if moving across any row involves adding a certain quantity to one entry to obtain the next and the same holds for moving down any column:

$$x_{i,j+1} = x_{i,j} + c,$$
 for some c and for all i,j .
 $x_{i+1,j} = x_{i,j} + d,$ for some d and for all i,j .¹³

This is a sufficient condition for additivity, not a necessary one; the definition of additivity is just that the interaction term (I) is zero and so the value within each cell is the sum of the grand mean and the two main effects.

Digitized by Arya Samaj Foundation Chennai and eGangotri

The case of two forces $(F_1 \text{ and } F_2)$ acting together, where their joint effect is simply the sum of what each would have achieved if it had acted alone, is just a special case of this formula:

	not-F ₁	F_1	
not-F ₂	0	c	
F_2	d	c+d	

The entries in an ANOVA table for nature and nurture might be additive in this sense, but this would not give local sense to the question of how much of Jane's height was due to her genes and how much to her environment. I conclude that additivity is neither necessary nor sufficient for locality.

The idea of height particles shows that the question of causal magnitude can be approached both locally and nonlocally with respect to the same physical system. Even if such particles existed, we still could pursue the analysis of variance to obtain a nonlocal answer to the question of causal magnitude. We also could do an intrinsic analysis of the relative contributions of Jane's environment and genes to her height. Indeed, we could obtain very different answers from these two procedures. If her genes produced 50 particles, but her environment only 10, we might judge from this intrinsic perspective that her genes contributed more. If, however, there is little genetic variation in Jane's population, the analysis of variance might conclude that environment made more of a difference than genes.

This shows why saying how much Jane's environment and genes contributed to her height can be a separate matter from saying whether her genes or environment made the greater difference. In the Newtonian case, these two questions are interchangeable and local (provided that counterfactuals are not allowed to backtrack). In the case of nature and nurture, they are different issues. How much difference Jane's environment and genes made to her height is answerable, though not locally. How much each contributed is an unanswerable question, since the empirical fact of the matter (I presume) is that there are no such things as height particles. Apportioning causal responsibility involves two issues, not just one.

¹⁴ One could take the position that the questions of how much each factor contributes and of how much difference each makes are equivalent and nonlocal in the bility should be apportioned. The fact that these questions come apart when we imagine a local determiner suggests, however, that the questions are better treated as separate ones from the start.

In the Newtonian case, the commensurability of gravity and electricity renders the question of causal contribution well-defined and local; in the nature/nurture case, we found neither commensurability nor locality. How much a causal factor contributes to its effect is. I would suggest, an inherently local question. But the question of how much difference a cause makes may be local or not, as is shown by the following example (suggested by Peter Woodruff).

Consider the fact that the distance traveled by a projectile shot from a cannon is influenced by both the muzzle velocity and the angle at which the gun is set. Suppose we fire a cannon and the shot goes half a mile. There is no saying how much each factor contributed to this outcome, nor which factor contributed more. The reason is that muzzle velocity and angle setting do not make their contributions in a common currency.

Yet, the question of what difference the two factors made may be locally or nonlocally assessible, depending on how we elaborate the story. Suppose the physical design of the gun shows that there are just two possible angle settings and two possible powder charges. If so, an intrinsic examination of the system allows us to fill out the relevant counterfactuals. Which factor made the larger difference is now a local matter.

On the other hand, suppose the gun has just one possible angle setting, though it may be packed with different amounts of powder. If the cannon maker designed several guns, each with its own fixed angle of fire, we would have to consider nonlocal facts to say what angle setting the gun would have had, if it had lacked the one it in fact had. Here the fixed setting is analogous to the fact that each of us has just one genetic endowment.

V. CONCLUSION

Causes are not necessary for their effects. It follows that one does not refute a causal claim by showing that the effect would have happened, even if the purported cause had not. Yet, when one shifts from the definition of causality to the problem of clarifying the concept of causal magnitude, something like the necessity thesis appears to be correct. If genes and environment are both causal factors influencing Jane's height, then genes have zero magnitude, provided that Jane would have had exactly the same height even if her genes had been different.

I therefore seem to find myself in the paradoxical position of saying that genes can be a cause of height, even if they are judged to have zero magnitude. But perhaps this air of paradox can be dispelled. It is not hard to fathom how causes can fail to be necessary for their effects. Suppose Watson would have shot Moriarty dead if Holmes had not. Holmes' pulling the trigger may have killed Moriarty, even though Holmes' firing was not necessary for Moriarty's death. The point may be put by saying that Holmes' firing, in this case, "made no difference." Causes may make no difference, but they are causes nonetheless.

I also have argued that the relative contribution of a cause and the difference it makes in the effect are separate issues. In the case of our Newtonian particle, this may look like a distinction without a difference. Disentangling the roles of genes and environment, however, shows this distinction to be a real one. Indeed, it is not just the science of ontogeny which draws this distinction. Holmes' pulling the trigger may have made no difference, but he certainly made a contribution to Moriarty's death.

A paradox that is less easy to dispel concerns the question of locality. Waiving as we have the quite separate issues raised by quantum mechanics, we may insist that causality is a local phenomenon. Holmes' firing causes Moriarty's death because there is a continuous process leading from one to the other. The causal relation obtains in virtue of this local circumstance. Watson's standing in the wings is relevant to the question of necessity, but not to the question of causality.

I have argued, however, that, even if causality is local, the magnitude of causality need not be. Jane's own genes and environment locally conspire to produce her adult height. But apportioning causal responsibility is not, in this case, a local matter. This is because it is a factual question, though not a local one, as to whether her genes or environment made the larger difference in her height. And it is not a factual matter at all, much less a local one, as to how much each contributed.

University of Wisconsin/Madison

ELLIOTT SOBER

PERSONAL IDENTITY AND SURVIVAL*

N this paper, I want to defend the view that, where there is personal survival, there must be personal identity. The kind of position which I oppose is that expressed by Derek Parfit in his 1976 footnote to the reprinted article "Personal Identity":

The real issue seems to me now this. Does personal identity just consist in bodily and psychological continuity, or is it a further fact, independent of the facts about these continuities? Our reactions to the 'problem cases' show, I believe, that we believe the latter. And we seem inclined to believe that this further fact is peculiarly deep and is all-or-nothing—we believe that in any describable case, it must hold completely or not at all. My main claim is *the denial of this further fact (ibid.*, fn. 37).

My main claim will be that the denial of this further fact is unwarranted. There is more to personal identity than the mere facts about bodily and psychological continuity. Even when there is full psychological continuity and connectedness between my present self and a future person, that future person need not be my survivor—indeed, my survival may depend on the existence of some other person.

To start, I need to consider the sort of puzzle cases which do seem to be possible and which may help to clarify our ideas about personal identity. I accept that, after transplanting my brain into someone else's brainless body, I would still exist. I would survive, but I would have a different body. This seems plausible because it draws support from the materialist theory which identifies mental states with brain states. Because the brain states have not altered, it would seem reasonable to suppose that the mental states are unchanged.

There is, though, a further thought experiment which has not been much considered. If my brain were to be decomposed and then reconstituted with every chemical replaced in exactly the same relative position, would I survive the operation? This thought experiment derives from a letter from Thomas Reid² to Lord Kames which includes the following passage:

^{*} I am very grateful for helpful comments on an earlier version of this paper by Jeffrey Gray, Paul Snowdon, and especially Kathy Wilkes. The paper was written while I was at Rutgers University and supported by the Fulbright Commission.

Philosophical Review, LXXX, 1 (Jan. 1971), reprinted in Jonathan Glover, ed., Philosophy of Mind (New York: Oxford, 1976). All page references are to the reprinted version.

² "Letter of 1775 to Lord Kames," in *Philosophical Works*, vol. 1 (Hildesheim: Olms, 1967).

replacement of just one of my neurons, but is it really conceivable that I should survive the whole of my brain being replaced by his? I think not. After all, at the end of the sequence of operations, all of his neurons would be in my skull in the very same order as they were before the operations started. We have, of course, achieved a brain transplant—only it has taken us a series of operations rather than a single one. It really does seem doubtful that I can conceive of myself surviving such a series of operations despite there being full psychological continuity between the person before and after the operations. On the contrary, the thought experiment seems to demonstrate clearly that there can be full psychological continuity between two different people.

That the person before and the person after the operation are quite distinct seems manifest when we consider a further twist to the story. If we imagine that my neurons were stored away somewhere as they were replaced, and then, at a later date were to be reconstructed in such a way as to constitute a structure exactly like that of my brain now, then we seem to have the Reid-derived puzzle case discussed earlier. If we reconstitute the neurons to form exactly the same brain structure, then it does seem that we do preserve the identity of the person who has undergone such deconstruction. It does seem that if, prior to the operation, I had been informed about its outcome, then I would have been concerned about the survival of the brain that was outside of my present body. After all, my identical twin is as distinct from me as is any other individual, and it does seem that it would be him occupying my body and not me. Full psychological continuity is not sufficient for survival. Personal identity cannot just consist in facts about bodily and psychological continuities but must also consist in some further fact.

III.

I have argued, in opposition to Parfit, that personal identity does not just consist in bodily and psychological continuity but seems to require some further fact. Parfit writes: "I cannot see how to disprove this . . . belief. I shall describe a problem case. But this can only make it seem implausible" ("Personal Identity," p. 143). How then do I resolve a puzzle case which supposes that my brain is divided into two identical parts, with each half being housed in a different body with the consequence that both the resulting persons have my character and apparent memories of my life. Parfit claims that there are only three possible descriptions of the result of the operation: (1) I do not survive; (2) I survive as one of the two people; (3) I survive as both. Parfit argues that neither (1) nor (2) are plausible descriptions, and settled for (3). About this he claims: "The relation of the original

person to each of the resulting people contains all that interests us—all that matters—in any ordinary case of survival. That is why we need a sense in which one person can survive as two" (*ibid.*, pp. 147/8).

I argued that the theoretical underpinning for the claim that I could survive as two people is deficient; full psychological continuity and connectedness does not preserve all that matters in my survival. There are then good reasons for rejecting all three descriptions of the possible results of Parfit's thought experiment. What should we conclude? Just this: Parfit's thought experiment, though imaginable, is not genuinely possible. Imaginability does not always lead to possibility. In this case, all three possible descriptions of the outcome of the Parfit-Wiggins split-brain operation are highly implausible, so it may well be more rational to deny the possibility of splitting, amoeba-like, into two than to accept the consequences of such a supposition. The kind of position I am adopting is a near neighbor of that expressed by David Wiggins⁵ in the following passage:

The conceptual possibility of a delta in the stream of consciousness jogs our whole focus on the concept of personhood. So, rather than jump to the conclusion that we have no idea at all what we are about, let us ask: is such a delta really a conceptual possibility? It would be easy to rule out such a thing in a lordly fashion . . . [b] ut what we really need is a good pretext to rule it out (ibid., pp. 157/8).

I would claim that the fact that the acceptance of the possibility of bifurcation leads one into a conceptual quagmire—committed to one of three highly implausible descriptions of the consequences of bifurcation—is sufficient pretext to rule out the supposed possibility as a genuine one. For those who, like Wiggins, do not believe this is sufficient ground, I shall bring forward some independent considerations which should lead us to reassess the coherence of Parfit's puzzle case.

The extremely powerful hold of Parfit's thought experiment on our imagination is largely due to Parfit's presentation of his philosophical speculations as an extrapolation from known facts. Those empirical facts, when brought into closer view, militate strongly against the coherence of Parfit's position and lend it no support whatsoever. Against the view that the person would not survive the operation, Parfit writes: "The trouble . . . is this. We agreed that I could survive if my brain were successfully transplanted. And people

⁵ "Locke, Butler and the Stream of Consciousness," in Amelie Rorty, ed., *The Identities of Persons* (Berkeley: California UP, 1976).

Digitized by Arya Samaj Foundation Chennai and eGangotri

have in fact survived with half their brains destroyed" ("Personal Identity," p. 144). The premise that "people have in fact survived with half their brain destroyed" is crucial to Parfit's argument. He continues: "It seems to follow that I could survive if half my brain were successfully transplanted and the other half were destroyed. But if this is so, how could I not survive if the other half were also successfully transplanted?" (ibid.).

The rhetorical question can be answered. Parfit seems to think that the two hemispheres (the two "half-brains") together constitute one whole brain. When Parfit writes that "people have in fact survived with half their brains destroyed," he seems to be referring to hemispherectomy patients (people who have had one of their hemispheres removed). Although such patients may lose either of their hemispheres and survive, they do, in every case, retain the brain stem. This is most important. If we were to transplant the removed hemisphere into someone else's brainless body, we would not create a second person. It is only if we were to transplant the hemisphere into a body that already had some brain in it, the brain stem, that there would be another person as a result. Once this feature of the puzzle case is highlighted, then it becomes quite clear that the relation between a surviving subject before and after the removal of one of the hemispheres is not duplicated in the case where a surviving subject's hemisphere is removed and placed in someone else's skull. In the former case, but not the latter, there is the fact of the persisting brain stem to consider and the psychological abilities and capacities that supervene on it.

We can now turn to Parfit's rejection of the description of the result of the imagined operation allowing only one person to survive. The argument he presents is brief: "The trouble here is that . . . each half of my brain is exactly similar, and so, to start with, is each resulting person. So how can I survive as only one of the two people? What can make me one of them rather than the other?" (ibid.). One answer is to question the wisdom of taking the imagined case seriously. The puzzle case is supposed to derive its philosophical importance and urgency from actual empirical facts—"people have in fact survived with half their brains destroyed"-but it is contrary to everything known about human brains to suppose that their two hemispheres are "exactly similar." If the puzzle case depends on this suggestion for its effect, then we can straightforwardly deny its applicability to human persons. On the other hand, if we were to suppose, correctly, that the two hemispheres (half-brains) were different, then it does seem that it could well be the presence of certain capacities, memories, etc., which makes me one person rather than the other.

Much of the philosophical speculation about the possibility of fission has been generated by the descriptions of behavior of commissurotomy patients. Parfit refers to these patients as actual instantiations of the possibility of a divided mind, apparently endorsing the surgeon's own description of the operation as the creations of "two separate streams of consciousness."

This description of the operation is controversial, and more recent experimental work conducted by Colwyn Trevarthen⁶ and his associates suggests that such a conclusion is incorrect. In his research on monkeys, Trevarthen found that the disconnected hemispheres were not wholly independent in their operations, for monkeys with their optic chiasma and interhemispheric commissures sectioned could make interocular comparisons of, for example, the relative size of two circles presented separately to each eye by means of polarizing filters. Following up this lead, Trevarthen later extended this line of research to humans. He found the by now familiar evidence of some form of disunity in the commissurotomy patient:

If a coloured photograph of an apple or a hammer or some other familiar object is projected at 15° to to the left of the point of visual fixation, the subject is unable to judge if it is the same or different from an object presented I5° right (ibid., p. 338).

Trevarthen also found evidence, however, that such subjects were, in fact, "able to perceive both halves of visual space at once, and to relate them together in a single perceptual synthesis" (ibid., p. 340). Trevarthen writes:

With 2 objects, for example, 5° black dots on the horizon at about 45° from the centre, one on each side, the subject judges accurately if they are being displaced together to the left or to the right, or in opposite directions so as to simultaneously approach or recede from the fixation point. If the dots undergo vertical displacements in steps of 10°, the subject judges if they remain together or move apart in opposite directions (ibid., p. 341).

This extremely important result (and others like it) seems extremely difficult to reconcile with the supposition that commissurotomy patients have two separate streams of conscious experiences. The patient makes a single judgment relating events that occur in both visual half fields and so it does seem reasonable to suppose that the

Man, Brain, Behaviour and Evolution, III (1970): 338-352.

^{6 &}quot;Two Mechanisms of Vision in Primates," Psychologische Forschung, XXXI (1968): 229–337; pp. 324–327 provide a summary of this research. "Experimental Evidence for a Brain Stem Contribution to Visual Perception in

Digitized by Arya Samaj Foundation Chennai and eGangotri

patient is aware of the events occurring in both half fields. Whereas the first of the experimental results quoted from Trevarthen indicates that the activities of the two hemispheres cannot always be integrated, the second shows that it is mistaken to conclude that the two hemispheres are independent entities supporting separate streams of consciousness. Since one can make such perceptual judgments only on the basis of one's own experiences, and since the commissurotomy patient is able to make some comparative judgments about events occurring in both the visual half fields, it is impossible to suppose that the two hemispheres contain two separate and independent streams of conscious experience.

The results of commissurotomy are contentious. The theorists who describe the operation as producing two streams of consciousness have the problem of explaining how the patients can integrate their activities in everyday life and also in the experimental conditions which have just been described. Those theorists, including myself, who believe there to be only one center of consciousness must explain the seemingly "self-contradictory" behavior of the patients manifested in experimental conditions. The position I advocate, but do not have space to argue here, is that the commissurotomy patient may have perceptual experiences located in each hemisphere but that these experiences are not conscious ones-nothing it is like to be having those experiences. The experimental evidence suggests that the unattended hemisphere supports nonconscious experiences (of which, in normal circumstances, a normal subject could ordinarily become conscious) and not an independent stream of consciousness. I am not the only theorist to have come to such a conclusion, for Marcel Kinsbourne⁸ writes:

While the intact brain is so organized that of competing responses to a single stimulus all but one are inhibited, in split-brain subjects two competing responses originating from the two sides may on rare occasions occur. However, it is not clear that both invoke attention; at least one may be automised. . . . So the person is at times informed by the logico-analytic control system of the left hemisphere and at times by the global-synthetic strategies characteristic of the right. But this does not imply concurrent streams of consciousness (*ibid.*, p. 287).

It is very interesting to notice that D. M. MacKay⁹ has also reached similar conclusions about the consciousness of commissurotomy pa-

^{8 &}quot;Cerebral Control and Mental Evolution," in Kinsbourne and W. L. Smith, eds., Hemisphere Disconnection and Cerebral Function (London: Thomas, 1974).

9 "Conscious Agency with Unsplit and Split Brains," in B. Josephson and V. Ramachandran, eds., Consciousness and the Physical World (London: Pergamon, 1980).

tients, even though his argument for that conclusion is very different. He writes that his

purpose . . . is . . . to show how unjustified it would be to assume that the peripheral splitting of the CNS, so as to produce independent sensori-motor and information-storage systems, is enough of itself to set up two independent "streams of consciousness." The key question, I am suggesting, is how many supervisory evaluators are able to function simultaneously. The answer for all of the human split-brain cases described so far appears to be: one. . . . At the outer levels of the evaluative hierarchy that are split, it is only to be expected that independent and even conflicting sub-criteria of evaluation can be established once conflicting patterns of experience have been mediated by the two hemisphere systems; but this does not show that either has broken free of the unifying central supervisory evaluation that (according to my conjecture) identifies the conscious agent (ibid., pp. 108/9).

It seems that we must regard the commissurotomy patient as a single subject of experience and action who has perceptual experiences localized in each of his two hemispheres. It does seem that we should suppose the unattended hemisphere to support nonconscious experiences and not an independent stream of conscious experiences. Trevarthen claims that the two hemispheres cannot be regarded as two separate and independent spheres of consciousness, because the brain stem plays an essential role both in the production of consciousness and in its unification, even in commissurotomy patients. Trevarthen's¹⁰ conclusions are unequivocal. In 1974 he wrote: "These two way links of hemispheres with the brain stem . . . make complete surgical duplication of consciousness in man an impossibility"(ibid., p. 261). There seem to be no empirical grounds for supposing that the kind of operation that Parfit wishes us to consider is indeed a genuine possibility. Indeed, the empirical evidence would seem to undermine the plausibility of the puzzle case.

I have tried to argue that there are no grounds for rejecting our beliefs in the existence of a further fact—over and above the facts about psychological continuity and connectedness—which is essential to our survival. I provided a science fiction thought experiment illustrating a case where there was a difference between exact similarity and identity, so that even in a case where Parfit would consider there to be full psychological continuity, there was a preferred candidate for our survival. I then addressed Parfit's argument which

Analysis of Cerebral Activities that Operate and Regulate Consciousness in Commissurotomy Patients," in S. Dimond and G. Beaumont, eds., Hemisphere Function in the Patients, 1974) Function in the Human Brain (New York: Wiley, 1974).

appeals to the imagined case where one divides: "The division of one stream of consciousness might be claimed to be deeply impossible. But what happens must be possible: and in the lives of several people this has happened" (*Reasons and Persons*, p. 278). The actual cases are not as Parfit imagines them to be—the counterintuitive consequences may be regarded as a reductio ad absurdum of the view that a division in the stream of consciousness is possible.

Parfit has a further argument against those who hold the combined view that any future person must be either me or someone else, and that there is always a deep difference between these two outcomes. Parfit considers the combined spectrum where, by tampering with my brain, there can be all of the possible degrees of either psychological or physical connectedness between me and some future person. About this Parfit argues:

If any future person must be either me or someone else, there must be a line in this range of cases up to which the resulting person would be me, and beyond which he would be someone else. If there is always a deep difference between someone's being me and his being someone else, there must always be a deep difference between the two of the cases in this spectrum. . . . These claims are false. There is no deep difference between any neighbouring cases in this range. The only differences are that, in one of the cases, the surgeons would replace a few more cells, and would make one more small psychological change (*ibid.*, p. 277).

The argument is invalid. By "replacing a few more cells," the surgeon is not necessarily making "one more small psychological change." Psychological properties supervene upon physiological properties—properties of the brain underly our psychological properties. Some changes in the brain will change our psychological properties, some may not. Of our psychological properties, some are essential properties of ourselves and some are contingent. Changes in cells which underly psychological properties that are only contingent properties of ourselves would not affect identity-changes in cells which underly our essential properties do affect our identity. Identification of these properties-identification of the facts over and above those about psychological continuity and connectedness which are necessary to explain survival—cannot be done independently of a theory about the nature, unity, and continuity of consciousness. An analysis of these further facts is long overdue in the philosophy of mind, and Parfit's arguments provide no reason for postponing this enterprise.

Medical Research Council/London

JOHN ROBINSON

BOOK REVIEWS

The Metaphysics of Modality. GRAEME FORBES. New York: Oxford University Press, 1985. 257 p. Cloth \$29.95.*

Since Naming and Necessity,1 metaphysics has returned to its old preoccupation with modality. In The Metaphysics of Modality, Graeme Forbes has consolidated much of the most important work of this period, and, for that alone, the book is well worth reading. Nearly every topic of recent interest receives thoughtful attention: propositional and quantified modal logic; the de re/de dicto distinction and de re skepticism; possible worlds semantics and possible worlds metaphysics; transworld identity; counterpart theory; the essential natures of sets, organisms, artifacts, substances, properties, and events; the ground of modality; and modal knowledge. But the real reason to read the book is that Forbes has imaginative and provocative things to offer on nearly all these topics. With regard to possible worlds, for example, he maintains that there are no such things, but that possible worlds semantics is not thereby rendered pointless. As to the ground of modality, he rediscovers and rehabilitates the old idea that necessity and possibility arise out of the operations of the mind.2 Identity and essence are the book's central themes, however, and here the argument is approximately as follows.

De re modal (temporal) attributions differ from de dicto in that their truth turns on "facts about transworld or transtemporal identity, facts to the effect that a certain individual at one world or time is identical to a certain individual at another" (50).3 But what if there are no facts of transworld identity? Then de re modal attributions must lack truth conditions (and so presumably meaning). And this is what the de re modal skeptic (represented here by W. V. O. Quine)

^{*} William Taschek and Lawrence Sklar helped me think through the issues of this

Saul Kripke (Cambridge: Harvard, 1980).

But, for second thoughts, see his "In Defense of Absolute Essentialism," Midwest Studies in Philosophy XI: Studies in Essentialism, Peter A. French, Theodore E. Uehling, Jr., and Howard K. Wettstein, eds. (Minneapolis: Minnesota UP, 1986), pp. 3-31, esp. section 3.

For results which tend to bear this out, see Kit Fine's "Model Theory for Modal Logic—I: The De Re/De Dicto Distinction," Journal of Philosophical Logic, VII (1978): 125–156. Notice that Forbes continues to frame the issues in terms of possible worlds even after his denial that they are real, probably on the theory that an elucidation of transworld identity can be regarded as an elucidation of the boundaries of possibility for ordinary things, an enterprise which an anti-realist about worlds must take seriously if he thinks there are objective modal facts, even if they are not about worlds" (96/7).

⁰⁰²²⁻³⁶²X/88/8506/0329\$00.90

maintains. Yet, unless transworld identity is less secure than transtemporal, the *de re* modal skeptic should also, in consistency, be a *de re* temporal skeptic, and so deny that "this was here yesterday" makes good sense. According to Quine, however, transworld identity is (at least prima facie) less secure, because "continuity of displacement, distortion, and chemical change" afford criteria of transtemporal, but not transworld, identity (53). Whence the challenge for *de re* believers is to show that there can also be criteria of transworld identity.

Why this challenge needs to be taken seriously is not altogether obvious. For initially, at least, it is uncertain whether continuity, or anything else, does give a criterion of transtemporal identity (as Forbes well realizes). More to the point, unless the well-definedness of transtemporal identity is on account of the availability of criteria, why should a lack of criteria in the transworld case be cause for concern? Some democracies have parliaments; but, unless their democratic character is known to depend on this, the democratic character of other societies does not fall under suspicion until it can be shown that they too have parliaments. Be that as it may, Forbes maintains that criteria of transworld identity are in any case available, in the form of individual essences. And his argument for this is the most interesting thing in the book.

What are individual essences? Simplifying slightly, a set *I* of properties is an essence of an individual thing *a* just in case possession of *I* is both necessary and sufficient for identity with *a*. Immediately on this characterization comes a problem, however: for every possible *a* is trivially possessed of the essence containing *identity with a* and nothing else. Trivial essences affording no criterion of transworld identity, substantial essences are required; but to specify what these are is not easy (let us assume that the notion is sufficiently well-defined). What Forbes proposes is that an essence is nontrivial if all its members are, where an essential property is trivial if (and here I cite only the clause most immediately relevant to the problem at hand) it follows from existence— or identity—properties "by logic alone" (99). But this seems unsatisfactory, for it leaves the door open to

What it is for x's possessing P to entail logically that it possesses R is not obvious.

⁴ Criteria may indeed assume a special importance on Quine's "selectionist" conception of transtemporal identity. According to Quine, "any two momentary objects, . . . taken at different moments, are time slices of some same time-extended object." That being so, "identification of an object from moment to moment by imposing criteria. [Quotations are from "Worlds Away," this JOURNAL, LXXIII, 570, and the property of the prope

properties like that of being a P identical with a (where P is any nontrivial essential property of a) which, though they generate individual essences, afford no criterion of transworld identity. Nor is the problem avoided by calling an essential property trivial if it logically entails existence— or identity—properties, for there will still be properties like that of being a P identical with a, or an R (R is any property not essential to a). Resolve these difficulties of characterization how we will, the real question remains: Can intuitively substantial essences be found? Inasmuch as the essences Forbes goes on to argue for are, by any reasonable standard, substantial, everything depends on the adequacy of the argument.

Two fundamental principles underly Forbes' case for essences. According to the principle of grounding, transworld identities (diversities) must always obtain in virtue of significantly other properties of the entity (entities) involved. According to the intrinsicness principle, the properties that ground an identity (diversity) are always intrinsic to the identical (diverse) entities. Neither principle is precisely formulated, but Forbes' examples afford an initial sense of their intended meaning.

Consider the supposition that things could have been exactly as they are except that the steel tower in Paris opposite the Palais de Chaillot is different from the one actually there. To make sense of this supposition, it is not permitted to imagine that the tower is made of different metal from the metal which actually constitutes it, or that it has a different design, or designer, or history. The *only* respect in which the imagined situation is to differ from the actual world is in the identity of the tower. The extent to which such a difference seems unintelligible is some measure of the plausibility of the view that transworld differences must be grounded (128).

Suppose . . . that A is an actual but untwinned human being. Could A have been an identical twin? . . . an affirmative answer to this question amounts to the claim that there is a possible world in which the zygote of A in that world undergoes fission . . . , followed by the non-standard separation giving rise to two individuals B and C; and furthermore, that

⁶ 'Intrinsic' does not mean nonrelational (there would be counterexamples to intrinsicness if it did). In one place, Forbes seems to suggest that it is sufficient for P to be extrinsic to a that: for some world w, whether a has P at w is sensitive to 'features of objects and events . . . causally isolated [from a] at w'' (141). But, in responding to the objection that an Episcopalian chapel can be distinct from a Buddhist monastery though they differ only extrinsically (different people own them), he offers that "the issue of who owns the building is not at all causally isolated from the issue of which denomination it houses" (142). Not from the issue, perhaps; but certainly from the alleged nonidenticals themselves! So it is not entirely clear what 'intrinsic' means. For a better formulation of the intrinsic/extrinsic distinction, see Forbes' "In Defense of Absolute Essentialism," op. cit.

A is identical to B or C. But we ought to be very reluctant to say that there is such a world. Of course, it is possible that A's zygote divides in the required way, but . . . it is evident that there is nothing in the situation to determine which of B and C is identical to A; nothing grounds one of the identities rather than the other (129).

Insofar as diversity is grounded, the actual and counterfactual towers, which share all conceivable grounding properties, are identical; insofar as identity is, A must be identical to both of B and C, or to neither (hence presumably to neither). Intrinsicness adds to this that the actual tower is identical to any intrinsic likeness of the counterfactual (even a tower beside the Palais du Chaillot), and that A is distinct from any intrinsic likeness of B (even if that individual is untwinned on account of the early disintegration of C's zygote).

At the risk of distorting a rough but workable maxim into a bloodless if precise postulate, let us interpret grounding as saying that identity supervenes on certain other properties, in the sense that there can be no difference in identity without a difference in these other properties (intrinsicness then adds to this that these other properties are intrinsic to their bearers). Specifically, given a, b, c, and d existing in worlds w_1 , w_2 , w_3 , and w_4 (respectively), if $G(a, w_1)$ and $G(b, w_2)$ are the same as $G(c, w_3)$ and $G(d, w_4)$ (respectively), then a = b if and only if c = d (here G(x, w) is the set of x's grounding properties in world w). To obtain the Eiffel Tower case, let a = b = c= the Eiffel Tower, $w_1 = w_2 = w_3$ = the actual world, and d = alikeness of the Eiffel Tower in w_4 . For the identical twin case, set a=c= the untwinned individual, b= the first twin, d= the second, w_1 = w_3 = the actual world, and w_2 = w_4 = the world of b and d. As for intrinsicness, it adds to grounding that only intrinsic properties of xcan go into G(x, w); the reader may verify that, thus interpreted, it has the consequences mentioned above.7

Though neither principle is obvious, let us concentrate on grounding (which intrinsicness anyway presupposes). Undeniably it pronounces rightly and convincingly on a wide range of cases. On the other hand,

consider a world w in which there are two qualitatively indiscernible iron globes which have always and will always exist; that it all there is to w. But

 7 Several things make me uneasy about interpreting grounding as a supervenience principle. First, Forbes treats grounding as made up of two distinct subprinciples, one about identity and the other about diversity; but this distinction seems to disappear on the supervenience formulation. Second, the supervenience formulation entails a version of the intraworld identity of indiscernibles, to which Forbes seems to regard himself as uncommitted (149). Third, it asks too much that a and b should have the *same* grounding properties as c and d; it ought to be enough that they are precisely as similar as c and d (the identical twins might be a case in point).

neither globe is essentially immortal, there are no restrictions on the times at which either globe could cease to exist. . . . Thus there are worlds u and v just like w, except that in u one of the globes ceases to exist at a time t . . . while in v it is the other globe which ceases to exist then. . . . In this set-up, . . . [perhaps] the facts about transworld identity are primitive, i.e., ungrounded. . . . (149-150).

In Forbes' view, the difficulty arises out of a wrong conception of possible worlds. Specifically, if we think of w "as a course of events and of u and v as courses of events 'branching' from w at the time t when one globe ceases to exist in u and the other in v," then "the transworld identities are explained by transtemporal identities across the branch-point at t" (150). What grounds the distinctness of the long-lived globe in v with the long-lived globe in v is that, "if we trace back in v from some point after v into v, and if we trace back in v from some point after v into v, we arrive at different globes' (150). In other words, worlds v and v share a history with v, and the long-lived globes in v and v are identical with distinct globes in v in virtue of their different relations to this common history.

Thus, branching appears to offer a mechanism whereby prima facie ungrounded transworld diversities may nevertheless be grounded in comparatively unproblematic intraworld diversities. True, the branching conception needs to be liberalized in light of other examples. For there are also distinct possible worlds r and s, in one of which only the first w-globe exists and in the other of which only the second does; and such worlds do not share a history with w. Still, they do have *some* overlap with w; they overlap with w on what Forbes calls a "separable" course of events (initial world-histories are a special case). And the previous solution generalizes, if we "count amongst worlds branching from w" worlds overlapping with w on any "separable course of events" (151).

Notwithstanding its considerable ingenuity, this solution raises awkward questions. When one world branches from another, is that in virtue of anything? If not, then it may seem as though we had traded ungrounded identity for a comparably baffling notion of ungrounded branching, or even, given the intimate connections between branching, overlap, and identity, readmitted ungrounded identity under another name. Yet, if branching relations are grounded, what are they grounded in? Not overlap, because overlap is partly a matter of the transworld identity for which branching was supposed to provide part of the ground. And not local indiscernibil-

⁸ The example comes from Robert Merrihew Adams, "Primitive Thisness and Primitive Identity," this JOURNAL, LXXVI, 1 (January 1979): 5–26.

ity either, because prima facie indiscernibility is insufficient for identity and hence insufficient for overlap, and genuine indiscernibility trades in the very branching relations for which a basis was sought. To see the problem, let w# be a world prima facie indiscernible from w, but containing different globes (a simple argument shows that w# exists). Now, although r and s branch from w, they cannot branch from w#, or else the r- and s-globes would be identical to the w#-globes. What can be responsible for this difference in branching? By hypothesis, w and w# are prima facie indiscernible, so it can only be (i) the differences in the identities of their inhabitants, or (ii) the difference we set out to explain. Yet, obviously neither (i) nor (ii) can be appealed to here, on pain of circularity. What this suggests is that some ungrounded relation may have to be tolerated in any case. Why not identity?

Even if it is allowed that whether worlds branch is not something that requires grounds, there is a more urgent question about how they branch. Worlds r and s are prima facie indiscernible, and neither branches from the other. So what grounds the transworld diversity of their globes? Presumably the answer is that r and s branch from at least one world (e.g., w) in relation to which their inhabitants may be distinguished. But this only pushes the question back a step. For what is there about r and s themselves to determine that they branch from a world like w in such a way as to make for the difference between their globes? That is, what is there about r and s to determine that they are built on distinct courses of events in some such world? What gives the question its urgency is that there are other, prima facie indiscernible, pairs of worlds (e.g., r and r) which, though they also branch from w, do it differently from r and s. The upshot is that there threaten to be bare facts not only about whether worlds branch, but also about how they branch. And these latter facts can be extremely similar to identity facts; or, to the extent that they are different, can seem to be grounded in identity facts rather than the other way around. For to say that the r- and s-globes are distinct because there are worlds (e.g., w) in which they (or their histories) stand apart, is arguably to get the intuitive order of explanation reversed: it is because the globes are distinct that they (or their histories) can stand apart.

Notwithstanding these theoretical difficulties, Forbes' applications of grounding and intrinsicness are sensible and even compelling. Thus, we should ask, granted grounding and intrinsicness, how well do they argue for individual essences? Take the case of organisms, specifically Forbes' argument that they have their origins es-

sentially (an organism's "origin" is the "antecedent entity or entities" from which it directly develops). Suppose per absurdum that they do not. Then tree T, which sprang from acorn A in the actual world $w^@$, does not come from A in some other world w. Suppose for the sake of argument that T comes from acorn A' in w; otherwise let T vary as little as possible between w and $w^@$. Now, this supposition can be brought into conflict with intrinsicness, as follows.

Most of us would agree that acorn A could have been planted at some small distance from where it actually was, in fact that there are various possible worlds in which this is what happens. Select from these worlds a world u that is as similar as possible to w (compatibly with the presence of this A-tree). In particular, u should contain a tree T_1 exactly like T is in w, in the sense that it grows in the same place, from the same acorn A', and has the same intrinsic properties throughout its existence (let the tree which grows from A in u be called T_2). Again, to most of us, it seems evident that the actual tree Tcould have been exactly like the tree that grows from A in u, i.e., the actual tree T could have grown from acorn A at a small distance from its actual location, and had the matter, shape, and so on of the tree that grows from A in u. Let v be a world in which this is what happens; thus, v is very much like u, except that in v there is no A'-tree, nor any tree at all in the vicinity of \hat{T} . The question is: What are the identity relations between T, T_1 , and T_2 ? Since T_2 is, in u, intrinsically indiscernible from T as it is in v, intrinsicness implies that $T = T_2$. Thus, T_1 , which is distinct from T_2 , must be distinct from T. On the other hand, T_1 , in u, has exactly the intrinsic properties that T has in w. So intrinsicness implies that T_1 is identical to T after all, contradicting our previous conclusion that it was not. So it was wrong to suppose that there was a world such as w, a world in which T did not spring from its actual acorn. Thus, it is essential to Tto have come from its actual acorn.

Naturally one wants some idea of this argument's general form. In broad outline, what is going on is perhaps this (I have taken some liberties). To prove that P is essential to a, begin by assuming that identity is intrinsic. Now, suppose that P and a are related as follows: necessarily, if anything x lacks P, it is possible that a duplicate of x should coexist with a duplicate of a as it might have been (put this by saying that -P extrudes a likeness of a). By intrinsicness, if -P extrudes a likeness of a, then, necessarily, anything which lacks P is liable to find itself accompanied by a. And this shows that, necessar-

 $^{^9}$ In keeping with his plan of uncovering not only essential properties but individual essences, Forbes goes on to argue that an organism's essence is its origin plus its kind (the argument will not be examined here).

Digitized by Arya Samaj Foundation Chennai and eGangotri

ily, a thing which lacks P is not a; which is just to say that P is essential to a.

Who might resist such a conclusion, and why? De re skeptics, certainly, simply because they find de re modal claims unintelligible; anti-essentialists, because they understand such claims well enough to have concluded that all but the most trivial are false; and, finally, dissident essentialists, who, though happy to accept some de re modal claims, happen to disagree with this one. Now, of the argument's two assumptions—the intrinsicness assumption and the extrusion assumption—which ought skeptics (anti-essentialists, dissident essentialists) to reject? Skeptics and anti-essentialists would be wise to reject both assumptions: intrinsicness, because (even by itself) it entails substantive de re modal theses; extrusion, because it is a substantive de re modal thesis. But what about the dissident essentialist, who (let us suppose) accepts intrinsicness, but wants to resist Forbes' essentialist conclusions?

Assuming intrinsicness, if -P extrudes a likeness of a, then a has P essentially. Thus, we need to know how extrusion theses are justified, and which extrusion theses are true. How is it that Forbes can show that an oak tree's acorn is essential to it, but not, say, its place or matter? The answer must be that, whereas originating from a different acorn does extrude a likeness of the tree, being in a different place (having different matter) does not. Spelling this out, only if not deriving from acorn A prevents a thing from monopolizing the materials for identity with T, is A shown to be essential to T. And, unless something can commandeer the resources for identity with T despite a shift in location or matter, location and matter are, contrary to expectation, shown to be essential to T.

Thus, our questions are: What establishes (1) that, unless something derives from A, it cannot commandeer the resources for identity with T? And (2) that something with a different location, or different matter, from T can monopolize the materials for identity with T? Unquestionably, our confidence in (1) rests at least partly on our confidence that properties monopolized by entities not deriving from A are always inessential to T. Granted, a tree not deriving from A can take over all of T's spatial career (its position through time), and also as much of its material career (its matter through time) as

¹⁰ At one point, Forbes suggests that a tree's location cannot be essential to it because location is not an intrinsic property of trees (144). But how do we know this? Admittedly, location is not a nonrelational property of trees, but this is not does not enter into any causal relation with the organism" (144), something similar tree's place is "causally isolated" from the *issue* of its transworld identity relations is just what is in question.

comports with the essential facts about acorn A. But neither of these is essential to T, and so their unavailability is no bar to the introduction of T on the side. If this is right, then, without assuming that various of its other properties are inessential to T, Forbes cannot prove that T essentially comes from A. Thus, it becomes important to know what blocks the argument that T essentially possesses aspects of its spatial or material career; and this takes us to (2). What shows that something which is never where T is in actuality, or nearly never composed of what T is in actuality composed of, might nevertheless monopolize the materials for being T? Apparently, the answer is that such a thing might nevertheless originate from acorn A; and so prevent anything else from coming from A; and thereby prevent anything else from being T. But this, of course, presupposes that originating from acorn A is essential to T. To put both halves of the story together, the argument that T's acorn is essential to it depends on the assumption that its location and matter are not; yet, they may be, and indeed there will be an argument that they are, unless we know that T has its acorn essentially.

Many philosophers would agree with Forbes that an oak tree's origin is essential to it, and that its location and matter (at any rate, most of its matter) are not. But the point remains that the essentialist thesis seems to depend on the inessentialist thesis, and the inessentialist thesis escapes probable refutation only if the essentialist thesis is true. Nothing prevents a dissident from denying the one, provided that she is also ready to deny the other. Thus, even if the dissident is wrong, Forbes' argument does not show that she is. And *some* dissident might, after all, be right.

None of this will come as news to Forbes, but it allows us to focus his accomplishment. Nobody should think that Forbes has established substantial essentialist conclusions on the basis of purely "structural" considerations. What he has done is to bring out hidden connections between superficially independent de re modal claims. Application of his methods may help essentialists to progress beyond brute conflicts of particular modal intuition to something more constructive and more hopeful, namely, the elaboration and comparison of competing essentialist systems. On what basis this larger competition is to be adjudicated is a hard question, but one on which Forbes makes an intriguing start. Not to begin another long story, let me end by urging metaphysicians to read this book. There is nothing better on modality.

STEPHEN YABLO

University of Michigan/Ann Arbor

¹¹ See ch. 9, "The Justification of Modal Concepts"; and "In Defense of Absolute Essentialism," pt. III.

NOTES AND NEWS

The editors report with sorrow the death of Roderick Firth, Alford Professor Emeritus of Philosophy at Harvard University, on December 22, 1987. Professor Firth received his S.B. degree from Haverford College in 1938 and Ph.D. degree from Harvard in 1943. Before joining the Harvard faculty in 1953, he taught at the College of William and Mary and at Swarthmore College. He was 70 years old at the time of his death. Donations in his memory may be made to the Society of Friends, 5 Longfellow Place, Cambridge, MA 02138.

Memphis State University will hold its annual Spindel Philosophy Conference on October 20–22, entitled "Aristotle's Ethics." Participants will include Julia Annas (Arizona), John Cooper (Princeton), Daniel Devereux (Virginia), Terence Irwin (Cornell), David Keyt (Washington), Richard Kraut (Illinois/Chicago), Timothy Roche (Memphis State), Susan Sauve (Harvard), Gisela Striker (Columbia), Nicholas White (Michigan), Steven White (Carleton), Jennifer Whiting (Pittsburgh), and Charles Young (Claremont). Further information may be obtained by contacting Timothy Roche, Memphis State Univ., Memphis, TN 38152.

CHARLES C THOMAS • PUBLISHER

ETHICAL DILEMMAS FOR ACADEMIC PROFESSIONALS edited by **Stephen L. Payne** and **Bruce H. Charnov.** This book vividly describes important ethical dilemmas faced by academic professionals, provides perspectives on confronting these dilemmas, and challenges readers to develop their own viewpoints. Chapters individually address such topics as fundamental means to ethical issues in the research publication process, the academician as a consultant, ethics in professional organizations, and faculty-administrative relationships. Case studies are provided to enhance experiential learning. '87, \$29.75

BEYOND DEATH: Evidence for Life After Death by Robert Almeder. This thought-provoking and readable review of the issue of life after death assesses the collective force, or intellectual merit, of the most convincing and best-documented cases currently available. The author examines the subject of reincarnation with discussions of memory studies and recognition, responsive xenoglossy, the case of Bishen Chand, the Shanti Devi case, and the skeptic's reply to these studies. Following chapters cover apparitions of the dead and the Watseka Wonder, out-of-body experiences, and communications from the dead. '87, \$24.50

Order direct for fastest results • Write or call (217) 789-8980 • Books sent on approval Postage paid on MasterCard, Visa & prepaid orders • Catalog sent on request

2600 South First Street • Springfield • Illinois • 62794-9265

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 7, JULY 1988

THE RETURN OF THE GENE*

Te have two images of natural selection. The orthodox story is told in terms of individuals. More organisms of any given kind are produced than can survive and reproduce to their full potential. Although these organisms are of a kind, they are not identical. Some of the differences among them make a difference to their prospects for survival or reproduction, and hence, on the average, to their actual reproduction. Some of the differences which are relevant to survival and reproduction are (at least partly) heritable. The result is evolution under natural selection, a process in which, barring complications, the average fitness of the organisms within a kind can be expected to increase with time.

There is an alternative story. Richard Dawkins1 claims that the "unit of selection" is the gene. By this he means not just that the result of selection is (almost always) an increase in frequency of some gene in the gene pool. That is uncontroversial. On Dawkins's conception, we should think of genes as differing with respect to properties that affect their abilities to leave copies of themselves. More genes appear in each generation than can copy themselves up to their full potential. Some of the differences among them make a

* We are equally responsible for this paper which was written when we discovered that we were writing it independently. We would like to thank those who have offered helpful suggestions to one or both of us, particularly Patrick Bateson, Robert Broad Lawrentin Lisa Lloyd Robert Brandon, Peter Godfrey-Smith, David Hull, Richard Lewontin, Lisa Lloyd, Philip Pettit, David Scheel, and Elliott Sober.

The claim is made in *The Selfish Gene* (New York: Oxford, 1976); and, in a somewhat modified form, in The Extended Phenotype (San Francisco: Freeman, 1989) West of the final section 1982). We shall discuss the difference between the two versions in the final section of this paper. of this paper, and our reconstruction will be primarily concerned with the later version of Deviated our reconstruction will be primarily concerned with the later version of Dawkins's thesis. We shall henceforth refer to The Selfish Gene as SG, and to The E and to The Extended Phenotype as EP. To forestall any possible confusion, our reconstructions are the provocative claims. reconstruction of Dawkins's position does not commit us to the provocative claims about altraice of SG fastened. about altruism and selfishness on which many early critics of SG fastened.

Digitized by Arya Samaj Foundation Chennai and eGangotri

difference to their prospects for successful copying and hence to the number of actual copies that appear in the next generation. Evolution under natural selection is thus a process in which, barring complication, the average ability of the genes in the gene pool to leave copies of themselves increases with time.

Dawkins's story can be formulated succinctly by introducing some of his terminology. Genes are *replicators* and selection is the struggle among *active germ-line* replicators. Replicators are entities that can be copied. Active replicators are those whose properties influence their chances of being copied. Germ-line replicators are those which have the potential to leave infinitely many descendants. Early in the history of life, coalitions of replicators began to construct *vehicles* through which they spread copies of themselves. Better replicators build better vehicles, and hence are copied more often. Derivatively, the vehicles associated with them become more common too. The orthodox story focuses on the successes of prominent vehicles—individual organisms. Dawkins claims to expose an underlying struggle among the replicators.

We believe that a lot of unnecessary dust has been kicked up in discussing the merits of the two stories. Philosophers have suggested that there are important connections to certain issues in the philosophy of science: reductionism, views on causation and natural kinds, the role of appeals to parsimony. We are unconvinced. Nor do we think that a willingness to talk about selection in Dawkinspeak brings any commitment to the adaptationist claims which Dawkins also holds. After all, adopting a particular perspective on selection is logically independent from claiming that selection is omnipresent in evolution.

In our judgment, the relative worth of the two images turns on two theoretical claims in evolutionary biology.

- 1. Candidate units of selection must have systematic causal consequences. If Xs are selected for, then X must have a systematic effect on its expected representation in future generations.
- 2. Dawkins's gene selectionism offers a more general theory of evolution. It can also handle those phenomena which are grist to the mill of individual selection, but there are evolutionary phenomena which fit the picture of individual selection ill or not at all, yet which can be accommodated naturally by the gene selection model.

Those skeptical of Dawkins's picture—in particular, Elliott Sober, Richard Lewontin, and Stephen Jay Gould—doubt whether genes can meet the condition demanded in (1). In their view, the phenomena of epigenesis and the extreme sensitivity of the phenotype to

gene combinations and environmental effects undercut genic selectionism. Although we believe that these critics have offered valuable insights into the character of sophisticated evolutionary modeling, we shall try to show that these insights do not conflict with Dawkins's story of the workings of natural selection. We shall endeavor to free the thesis of genic selectionism from some of the troublesome excresences which have attached themselves to an interesting story.

I. GENE SELECTION AND BEAN-BAG GENETICS

Sober and Lewontin² argue against the thesis that all selection is genic selection by contending that many instances of selection do not involve selection for properties of individual alleles. Stated rather loosely, the claim is that, in some populations, properties of individual alleles are not positive causal factors in the survival and reproductive success of the relevant organisms. Instead of simply resting this claim on an appeal to our intuitive ideas about causality, Sober has recently provided an account of causal discourse which is intended to yield the conclusion he favors, thus rebutting the proposals of those (like Dawkins) who think that properties of individual alleles can be causally efficacious.3

The general problem arises because replicators (genes) combine to build vehicles (organisms) and the effect of a gene is critically dependent on the company it keeps. However, recognizing the general problem, Dawkins seeks to disentangle the various contributions of the members of the coalition of replicators (the genome). To this end, he offers an analogy with a process of competition among rowers for seats in a boat. The coach may scrutinize the relative times of different teams but the competition can be analyzed by investigating the contributions of individual rowers in different contexts (SG 40/1 91/2, EP 239).

Sober's Case. At the general level, we are left trading general intuitions and persuasive analogies. But Sober (and, earlier, Sober and Lewontin) attempted to clarify the case through a particular example. Sober argues that heterozygote superiority is a phenomenon that cannot be understood from Dawkins's standpoint. We shall discuss Sober's example in detail; our strategy is as follows. We first set out Sober's case: heterozygote superiority cannot be understood as a gene-level phenomenon, because only pairs of genes can be, or fail to be, heterozygous. Yet being heterozygous can be causally

² "Artifact, Cause and Genic Selection," *Philosophy of Science*, XLIX (1982): 7-180

See Sober, The Nature of Selection (Cambridge: MIT, 1984), chs. 7-9, espeally 309-214 cially 302-314. We shall henceforth refer to this book as NS.

salient in the selective process. Against Sober, we first offer an analogy to show that there must be something wrong with his line of thought: from the gene's eye view, heterozygote superiority is an instance of a standard selective phenomenon, namely *frequency-dependent* selection. The advantage (or disadvantage) of a trait can depend on the frequency of that trait in other members of the relevant population.

Having claimed that there is something wrong with Sober's argument, we then try to say what is wrong. We identify two principles on which the reasoning depends. First is a general claim about causal uniformity. Sober thinks that there can be selection for a property only if that property has a positive uniform effect on reproductive success. Second, and more specifically, in cases where the heterozygote is fitter, the individuals have no uniform causal effect. We shall try to undermine both principles, but the bulk of our criticism will be directed against the first.

Heterozygote superiority occurs when a heterozygote (with genotype Aa, say) is fitter than either homozygote (AA or aa). The classic example is human sickle-cell anemia: homozygotes for the normal allele in African populations produce functional hemoglobin but are vulnerable to malaria, homozygotes for the mutant ("sickling") allele suffer anemia (usually fatal), and heterozygotes avoid anemia while also having resistance to malaria. The effect of each allele varies with context, and the contexts across which variation occurs are causally relevant. Sober writes:

In this case, the a allele does not have a unique causal role. Whether the gene a will be a positive or a negative causal factor in the survival and reproductive success of an organism depends on the genetic context. If it is placed next to a copy of A, a will mean an increase in fitness. If it is placed next to a copy of itself, the gene will mean a decrement in fitness (NS 303).

The argument against Dawkins expressed here seems to come in two parts. Sober relies on the principle

(A) There is selection for property P only if in all causally relevant background conditions P has a positive effect on survival and reproduction.

He also adduces a claim about the particular case of heterozygote superiority.

(B) Although we can understand the situation by noting that the heterozygote has a uniform effect on survival and reproduction, the property

Digitized by Arya Samaj Foundation Chennai and eGangotri

of having the A allele and the property of having the a allele cannot be seen as having uniform effects on survival and reproduction.

We shall argue that both (A) and (B) are problematic.

Let us start with the obvious reply to Sober's argument. It seems that the heterozygote superiority case is akin to a familiar type of frequency-dependent selection. If the population consists just of AAs and a mutation arises, the a-allele, then, initially a is favored by selection. Even though it is very bad to be aa, a alleles are initially likely to turn up in the company of A alleles. So they are likely to spread, and, as they spread, they find themselves alongside other a alleles, with the consequence that selection tells against them. The scenario is very similar to a story we might tell about interactions among individual organisms. If some animals resolve conflicts by playing hawk and others play dove, then, if a population is initially composed of hawks (and if the costs of bloody battle outweigh the benefits of gaining a single resource), doves will initially be favored by selection. 4 For they will typically interact with hawks, and, despite the fact that their expected gains from these interactions are zero, they will still fare better than their rivals whose expected gains from interactions are negative. But, as doves spread in the population, hawks will meet them more frequently, with the result that the expected payoffs to hawks from interactions will increase. Because they increase more rapidly than the expected payoffs to the doves, there will be a point at which hawks become favored by selection, so that the incursion of doves into the population is halted.

We believe that the analogy between the case of heterozygote superiority and the hawk-dove case reveals that there is something troublesome about Sober's argument. The challenge is to say exactly

what has gone wrong.

Causal Uniformity. Start with principle (A). Sober conceives of selection as a force, and he is concerned to make plain the effects of component forces in situations where different forces combine. Thus, he invites us to think of the heterozygote superiority case by analogy with situations in which a physical object remains at rest because equal and opposite forces are exerted on it. Considering the situation only in terms of net forces will conceal the causal structure of the situation. Hence, Sober concludes, our ideas about units of selection should penetrate beyond what occurs on the average, and

⁴ For details, see John Maynard Smith, Evolution and the Theory of Games (New York; Cambridge, 1982); and, for a capsule presentation, Philip Kitcher, Vaulting 1985), pp. 88-97.

Digitized by Arya Samaj Foundation Chennai and eGangotri

we should attempt to isolate those properties which positively affect survival and reproduction in every causally relevant context.

Although Sober rejects determinism, principle (A) seems to hanker after something like the uniform association of effects with causes that deterministic accounts of causality provide. We believe that the principle cannot be satisfied without doing violence to ordinary ways of thinking about natural selection, and, once the violence has been exposed, it is not obvious that there is any way to reconstruct ideas about selection that will fit Sober's requirement.

Consider the example of natural selection, the case of industrial melanism.5 We are inclined to say that the moths in a Cheshire wood. where lichens on many trees have been destroyed by industrial pollutants, have been subjected to selection pressure and that there has been selection for the property of being melanic. But a moment's reflection should reveal that this description is at odds with Sober's principle. For the wood is divisible into patches, among which are clumps of trees that have been shielded from the effects of industrialization. Moths who spend most of their lives in these areas are at a disadvantage if they are melanic. Hence, in the population comprising all the moths in the wood, there is no uniform effect on survival and reproduction: in some causally relevant contexts (for moths who have the property of living in regions where most of the trees are contaminated), the trait of being melanic has a positive effect on survival and reproduction, but there are other contexts in which the effect of the trait is negative.

The obvious way to defend principle (A) is to split the population into subpopulations and identify different selection processes as operative in different subgroups. This is a revisionary proposal, for our usual approach to examples of industrial melanism is to take a coarse-grained perspective on the environments, regarding the existence of isolated clumps of uncontaminated trees as a perturbation of the overall selective process. Nonetheless, we might be led to make the revision, not in the interest of honoring a philosophical prejudice, but simply because our general views about selection are consonant with principle (A), so that the reform would bring our treatment of examples into line with our most fundamental beliefs about selection.

In our judgment, a defense of this kind fails for two connected reasons. First, the process of splitting populations may have to continue much further—perhaps even to the extent that we ultimately

⁵ The locus classicus for discussion of this example is H.B.D. Kettlewell, *The Evolution of Melanism* (New York: Oxford, 1973).

conceive of individual organisms as making up populations in which a particular type of selection occurs. For, even in contaminated patches, there may be variations in the camouflaging properties of the tree trunks and these variations may combine with propensities of the moths to cause local disadvantages for melanic moths. Second, as many writers have emphasized, evolutionary theory is a statistical theory, not only in its recognition of drift as a factor in evolution but also in its use of fitness coefficients to represent the expected survivorship and reproductive success of organisms. The envisaged splitting of populations to discover some partition in which principle (A) can be maintained is at odds with the strategy of abstracting from the thousand natural shocks that organisms in natural populations are heir to. In principle, we could relate the biography of each organism in the population, explaining in full detail how it developed, reproduced, and survived, just as we could track the motion of each molecule of a sample of gas. But evolutionary theory, like statistical mechanics, has no use for such a fine grain of description: the aim is to make clear the central tendencies in the history of evolving populations, and, to this end, the strategy of averaging, which Sober decries, is entirely appropriate. We conclude that there is no basis for any revision that would eliminate those descriptions which run counter to principle (A).

At this point, we can respond to the complaints about the gene's eye view representation of cases of heterozygote superiority. Just as we can give sense to the idea that the trait of being melanic has a unique environment-dependent effect on survival and reproduction, so too we can explicate the view that a property of alleles, to wit, the property of directing the formation of a particular kind of hemoglobin, has a unique environment-dependent effect on survival and reproduction. The alleles form parts of one another's environments, and, in an environment in which a copy of the A allele is present, the typical trait of the S allele (namely, directing the formation of deviant hemoglobin) will usually have a positive effect on the chances that copies of that allele will be left in the next generation. (Notice that the effect will not be invariable, for there are other parts of the genomic environment which could wreak havoc with it). If someone protests that the incorporation of alleles as themselves part of the environment is suspect, then the immediate rejoinder is that, in cases of behavioral interactions, we are compelled to treat organisms as parts of one another's environments. The effects of playing hawk

⁶ In the spirit of Sober's original argument, one might press further. Genic selectionists contend that an A allele can find itself in two different environments, one in which the effect of directing the formation of a normal globin chain is

depend on the nature of the environment, specifically on the fre-

quency of doves in the vicinity.7

The Causal Powers of Alleles. We have tried to develop our complaints about principle (A) into a positive account of how cases of heterozygote superiority might look from the gene's eye view. We now want to focus more briefly on (B). Is it impossible to reinterpret the examples of heterozygote superiority so as to ascribe uniform effects on survival and reproduction to allelic properties? The first point to note is that Sober's approach formulates the Dawkinsian point of view in the wrong way: the emphasis should be on the effects of properties of alleles, not on allelic properties of organisms (like the property of having an A allele) and the accounting ought to be done in terms of allele copies. Second, although we argued above that the strategy of splitting populations was at odds with the character of evolutionary theory, it is worth noting that the same strategy will be available in the heterozygote superiority case.

positive and one in which that effect is negative. Should we not be alarmed by the fact that the distribution of environments in which alleles are selected is itself a function of the frequency of the alleles whose selection we are following? No. The phenomenon is thoroughly familiar from studies of behavioral interactions—in the hawk-dove case we treat the frequency of hawks both as the variable we are tracking and as a facet of the environment in which selection occurs. Maynard Smith makes the parallel fully explicit in his paper "How To Model Evolution," in John Dupre, ed., The Latest on the Best: Essays on Optimality and Evolution (Cambridge:

MIT, 1987), pp. 119-131, especially pp. 125/6.

⁷ Moreover, we can explicitly recognize the co-evolution of alleles with allelic environments. A fully detailed general approach to population genetics from the Dawkinsian point of view will involve equations that represent the functional dependence of the distribution of environments on the frequency of alleles, and equations that represent the fitnesses of individual alleles in different environments. In fact, this is just another way of looking at the standard population genetics equations. Instead of thinking of W_{AA} as the expected contribution to survival and reproduction of (an organism with) an allelic pair, we think of it as the expected contribution of copies of itself of the allele A in environment A. We now see W_{AS} as the expected contribution of A in environment S and also as the expected contribution of S in environment A. The frequencies p, q are not only the frequencies of the alleles, but also the frequencies with which certain environments occur. The standard definitions of the overall (net) fitnesses of the alleles are obtained by weighting the fitnesses in the different environments by the frequencies with which the environments occur.

Lewontin has suggested to us that problems may arise with this scheme of interpretation if the population should suddenly start to reproduce asexually. But this hypothetical change could be handled from the genic point of view by recognizing an alteration of the coevolutionary process between alleles and their environments: whereas certain alleles used to have descendants that would encounter a variety of environments, their descendants are now found only in one allelic environment. Once the algebra has been formulated, it is relatively straightforward to extend the reinterpretation to this case.

Consider the following division of the original population: let P. be the collection of all those allele copies which occur next to an S allele, and let P_2 consist of all those allele copies which occur next to an A allele. Then the property of being A (or of directing the production of normal hemoglobin) has a positive effect on the production of copies in the next generation in P_1 , and conversely in P_2 . In this way, we are able to partition the population and to achieve a Dawkinsian redescription that meets Sober's principle (A)—just in the way that we might try to do so if we wanted to satisfy (A) in understanding the operation of selection on melanism in a Cheshire wood or on fighting strategies in a population containing a mixture of hawks and doves.

Objection: the "populations" just defined are highly unnatural, and this can be seen once we recognize that, in some cases, allele copies in the same organisms (the heterozygotes) belong to different "populations." Reply: so what? From the allele's point of view, the copy next door is just a critical part of the environment. The populations P_1 and P_2 simply pick out the alleles that share the same environment. There would be an analogous partition of a population of competing organisms which occurred locally in pairs such that some organisms played dove and some hawk. (Here, mixed pairs would correspond to heterozygotes).

So the genic picture survives an important initial challenge. The moral of our story so far is that the picture must be applied consistently. Just as paradoxical conclusions will result if one offers a partial translation of geometry into arithmetic, it is possible to generate perplexities by failing to recognize that the Dawkinsian Weltanschauung leads to new conceptions of environment and of population. We now turn to a different worry, the objection that genes are not "visible" to selection.

II. EPIGENESIS AND VISIBILITY

In a lucid discussion of Dawkins's early views, Gould claims to find a "fatal flaw" in the genic approach to selection. According to Gould, Dawkins is unable to give genes "direct visibility to natural selection."8 Bodies must play intermediary roles in the process of selection, and, since the properties of genes do not map in one-one fashion onto the properties of bodies, we cannot attribute selective advantages to individual alleles. We believe that Gould's concerns raise two important kinds of issues for the genic picture: (i) Can

^{8 &}quot;Caring Groups and Selfish Genes," in *The Panda's Thumb* (New York: Nor-1980) and Selfish Genes," in *The Panda's Thumb* (New York: NS 227) ton, 1980), p. 90. There is a valuable discussion of Gould's claims in Sober, NS 227 ff.

Dawkins sensibly talk of the effect of an individual allele on its expected copying frequency? (ii) Can Dawkins meet the charge that it is the phenotype that makes the difference to the copying of the underlying alleles, so that, whatever the causal basis of an advantageous trait, the associated allele copies will have enhanced chances of being replicated? We shall take up these questions in order.

Do Alleles Have Effects? Dawkins and Gould agree on the facts of embryology which subvert the simple Mendelian association of one gene with one character. But the salience of these facts to the debate is up for grabs. Dawkins regards Gould as conflating the demands of embryology with the demands of the theory of evolution. While genes' effects blend in embryological development, and while they have phenotypic effects only in concert with their gene-mates, genes "do not blend as they replicate and recombine down the generations. It is this that matters for the geneticist, and it is also this that matters for the student of units of selection" (EP 117).

Is Dawkins right? Chapter 2 of EP is an explicit defense of the meaningfulness of talk of "genes for" indefinitely complex morphological and behavioral traits. In this, we believe, Dawkins is faithful to the practice of classical geneticists. Consider the vast number of loci in *Drosophila melanogaster* which are labeled for eye-color traits—white, eosin, vermilion, raspberry, and so forth. Nobody who subscribes to this practice of labeling believes that a pair of appropriately chosen stretches of DNA, cultured in splendid isolation, would produce a detached eye of the pertinent color. Rather, the intent is to indicate the effect that certain changes at a locus would make against the background of the rest of the genome.

Dawkins's project here is important not just in conforming to traditions of nomenclature. Remember: Dawkins needs to show that we can sensibly speak of alleles having (environment-sensitive) effects, effects in virtue of which they are selected for or selected against. If we can talk of a gene for X, where X is a selectively important phenotypic characteristic, we can sensibly talk of the effect of an allele on its expected copying frequency, even if the effects are always indirect, via the characteristics of some vehicle.

What follows is a rather technical reconstruction of the relevant notion. The precision is needed to allow for the extreme environmental sensitivity of allelic causation. But the intuitive idea is simple: we can speak of genes for X if substitutions on a chromosome would lead, in the relevant environments, to a difference in the X-ishness of the phenotype.

Consider a species S and an arbitrary locus L in the genome of members of S. We want to give sense to the locution 'L is a locus

affecting P' and derivatively to the phrase 'G is a gene for $P^{*'}$ (where, typically, P will be a determinable and P^* a determinate form of P). Start by taking an environment for a locus to be an aggregate of DNA segments that would complement L to form the genome of a member of S together with a set of extra-organismic factors (those aspects of the world external to the organism which we would normally count as part of the organism's environment). Let a set of variants for L be any collection of DNA segments, none of which is debarred, on physico-chemical grounds, from occupying L. (This is obviously a very weak constraint, intended only to rule out those segments which are too long or which have peculiar physico-chemical properties). Now, we say that L is a locus affecting P in S relative to an environment E and a set of variants V just in case there are segments s, s^* , and s^{**} in V such that the substitution of s^{**} for s^* in an organism having s and s* at L would cause a difference in the form of P, against the background of E. In other words, given the environment E, organisms who are ss^* at L differ in the form of P from organisms who are ss^{**} at L and the cause of the difference is the presence of s^* rather than s^{**} . (A minor clarification: while s^* and s^{**} are distinct, we do not assume that they are both different from s.)

L is a locus affecting P in S just in case L is a locus affecting P in Srelative to any standard environment and a feasible set of variants. Intuitively, the geneticist's practice of labeling loci focuses on the "typical" character of the complementary part of the genome in the species, the "usual" extra-organismic environment, and the variant DNA segments which have arisen in the past by mutation or which "are likely to arise" by mutation. Can these vague ideas about standard conditions be made more precise? We think so. Consider first the genomic part of the environment. There will be numerous alternative combinations of genes at the loci other than L present in the species S. Given most of these gene combinations, we expect modifications at L to produce modifications in the form of P. But there are likely to be some exceptions, cases in which the presence of a rare allele at another locus or a rare combination of alleles produces a phenotypic effect that dominates any effect on P. We can either dismiss the exceptional cases as nonstandard because they are infrequent or we can give a more refined analysis, proposing that each of the nonstandard cases involves either (a) a rare allele at a locus L' or (b) a rare combination of alleles at loci L', L''. . . such that that locus (a) or those loci jointly (b) affect some phenotypic trait Q that dominates P in the sense that there are modifications of Q which prevent the expression of any modifications of P. As a concrete example, consider the fact that there are modifications at some loci in Drosophila which produce embryos that fail to develop heads; given such modifications elsewhere in the genome, alleles affecting eye color do not produce their standard effects!

We can approach standard extra-genomic environments in the same way. If L affects the form of P in organisms with a typical gene complement, except for those organisms which encounter certain rare combinations of external factors, then we may count those combinations as nonstandard simply because of their infrequency. Alternatively, we may allow rare combinations of external factors to count provided that they do not produce some gross interference with the organism's development, and we can render the last notion more precise by taking nonstandard environments to be those in which the population mean fitness of organisms in S would be reduced by some arbitrarily chosen factor (say, $\frac{1}{2}$).

Finally, the feasible variants are those which actually occur at L in members of S, together with those which have occurred at L in past members of S and those which are easily attainable from segments that actually occur at L in members of S by means of insertion, deletion, substitution, or transposition. Here the criteria for ease of attainment are given by the details of molecular biology. If an allele is prevalent at L in S, then modifications at sites where the molecular structure favors insertions, deletions, substitutions, or transpositions (so-called "hot spots") should count as easily attainable even if some of these modifications do not actually occur.

Obviously, these concepts of "standard conditions" could be articulated in more detail, and we believe that it is possible to generate a variety of explications, agreeing on the core of central cases but adjusting the boundaries of the concepts in different ways. If we now assess the labeling practices of geneticists, we expect to find that virtually all of their claims about loci affecting a phenotypic trait are sanctioned by all of the explications. Thus, the challenge that there is no way to honor the facts of epigenesis while speaking of loci that affect certain traits would be turned back

Once we have come this far, it is easy to take the final step. An allele A at a locus L in a species S is for the trait P^* (assumed to be a determinate form of the determinable characteristic P) relative to a local allele B and an environment E just in case (a) L affects the form of P in S, (b) E is a standard environment, and (c) in E organisms that are AB have phenotype P^* . The relativization to a local allele is necessary, of course, because, when we focus on a target allele rather than a locus, we have to extend the notion of the environment—as we saw in the last section, corresponding alleles are potentially important parts of one another's environments. If we say that A is for

 P^* (period), we are claiming that A is for P^* relative to standard environments and common local alleles or that A is for P^* relative to standard environments and itself.

Now, let us return to Dawkins and to the apparently outré claim that we can talk about genes for reading. Reading is an extraordinarily complex behavior pattern and surely no adaptation. Further, many genes must be present and the extra-organismic environment must be right for a human being to be able to acquire the ability to read. Dyslexia might result from the substitution of an unusual mutant allele at one of the loci, however. Given our account, it will be correct to say that the mutant allele is a gene for dyslexia and also that the more typical alleles at the locus are alleles for reading. Moreover, if the locus also affects some other (determinable) trait, say, the capacity to factor numbers into primes, then it may turn out that the mutant allele is also an allele for rapid factorization skill and that the typical allele is an allele for factorization disability. To say that A is an allele for P^* does not preclude saying that A is an allele for Q*, nor does it commit us to supposing that the phenotypic properties in question are either both skills or both disabilities. Finally, because substitutions at many loci may produce (possibly different types of) dyslexia, there may be many genes for dyslexia and many genes for reading. Our reconstruction of the geneticists' idiom, the idiom which Dawkins wants to use, is innocent of any Mendelian theses about one-one mappings between genes and phenotypic traits.

Visibility. So we can defend Dawkins's thesis that alleles have properties that influence their chances of leaving copies in later generations by suggesting that, in concert with their environments (including their genetic environments), those alleles cause the presence of certain properties in vehicles (such as organisms) and that the properties of the vehicles are causally relevant to the spreading of copies of the alleles. But our answer to question (i) leads naturally to concerns about question (ii). Granting that an allele is for a phenotypic trait P^* and that the presence of P^* rather than alternative forms of the determinable trait P enhances the chances that an organism will survive and reproduce and thus transmit copies of the underlying allele, is it not P^* and its competition which are directly involved in the selection process? What selection "sees" are the phenotypic properties. When this vague, but suggestive, line of thought has been made precise, we think that there is an adequate Dawkinsian reply to it.

The idea that selection acts directly on phenotypes, expressed in metaphorical terms by Gould (and earlier by Ernst Mayr), has been

Digitized by Arya Samaj Foundation Chennai and eGangotri

explored in an interesting essay by Robert Brandon.9 Brandon proposes that phenotypic traits screen off genotypic traits (in the sense of Wesley Salmon 10):

$$Pr(O_n/G\&P) = Pr(O_n/P) \neq Pr(O_n/G)$$

where $Pr(O_n/G\&P)$ is the probability that an organism will produce n offspring given that it has both a phenotypic trait and the usual genetic basis for that trait, $Pr(O_n/P)$ is the probability that an organism will produce n offspring given that it has the phenotypic trait. and $Pr(O_n/G)$ is the probability that it will produce n offspring given that it has the usual genetic basis. So fitness seems to vary more directly with the phenotype and less directly with the underlying

genotype.

Why is this? The root idea is that the successful phenotype may occur in the presence of the wrong allele as a result of judicious tampering, and, conversely, the typical effect of a "good" allele may be subverted. If we treat moth larvae with appropriate injections, we can produce pseudomelanics that have the allele which normally gives rise to the speckled form and we can produce moths, foiled melanics, that carry the allele for melanin in which the developmental pathway to the emergence of black wings is blocked. The pseudomelanics will enjoy enhanced reproductive success in polluted woods and the foiled melanics will be at a disadvantage. Recognizing this type of possibility, Brandon concludes that selection acts at the level of the phenotype.11

⁹ Gould, op.cit.; Mayr, Animal Species and Evolution (Cambridge: Harvard, 1963), p. 184; and Brandon, "The Levels of Selection," in Brandon and Richard Burian, eds., Genes, Organisms, Populations (Cambridge: MIT, 1984), pp.

Brandon refers to Salmon's "Statistical Explanation," in Salmon, ed., Statistical Explanation and Statistical Relevance (Pittsburgh: University Press, 1971). It is now widely agreed that statistical relevance misses some distinctions which are important in explicating causal relevance. See, for example, Nancy Cartwright, "Causal Laws and Effective Strategies," Noûs, XIII (1979): 419-437; Sober, NS ch 8; and Salmon, Scientific Explanation and the Causal Structure of the World

(Princeton: University Press, 1984).

Unless the treatments are repeated in each generation, the presence of the genetic basis for melanic coloration will be correlated with an increased frequency of grandoffspring, or of great-grandoffspring, or of descendants in some further generation. Thus, analogs of Brandon's probabilistic relations will hold only if the progeny of foiled melanics are treated so as to become foiled melanics, and the progeny of pseudomelanics are treated so as to become pseudomelanics. This point reinforces the claims about the relativization to the environment that we make below. Brandon has suggested to us in correspondence that now his preferred strategy for tackling issues of the units of selection would be to formulate a principle for identifying genuine environments.

Once again, there is no dispute about the facts. But our earlier discussion of epigenesis should reveal how genic selectionists will want to tell a different story. The interfering conditions that affect the phenotype of the vehicle are understood as parts of the allelic environment. In effect, Brandon, Gould, and Mayr contend that, in a polluted wood, there is selection for being dark colored rather than for the allelic property of directing the production of melanin, because it would be possible to have the reproductive advantage associated with the phenotype without having the allele (and conversely it would be possible to lack the advantage while possessing the allele). Champions of the gene's eye view will maintain that tampering with the phenotype reverses the typical effect of an allele by changing the environment. For these cases involve modification of the allelic environment and give rise to new selection processes in which allelic properties currently in favor prove detrimental. The fact that selection goes differently in the two environments is no more relevant than the fact that selection for melanic coloration may go differently in Cheshire and in Dorset.

If we do not relativize to a fixed environment, then Brandon's claims about screening off will not generally be true.12 We suppose

12 Intuitively, this will be because Brandon's identities depend on there being no correlation between O_n and G in any environment, except through the property P. Thus, ironically, the screening-off relations only obtain under the assumptions of simple bean-bag genetics! Sober seems to appreciate this point in a cryptic footnote (NS 229-230).

To see how it applies in detail, imagine that we have more than one environment and that the reproductive advantages of melanic coloration differ in the different environments. Specifically, suppose that E_1 contains m_1 organisms that have P(melanic coloration) and G (the normal genetic basis of melanic coloration), that E_2 contains m_2 organisms that have P and G, and that the probabilities $\Pr(O_n/G\&P\&E_1)$ and $\Pr(O_n/G\&P\&E_2)$ are different. Then, if we do not relativize to environments, we shall compute $Pr(O_n/G\&P)$ as a weighted average of the probabilities relative to the

$$\Pr(O_n/G\&P) = \Pr(E_1/G\&P) \cdot \Pr(O_n/G\&P\&E_1) + \Pr(E_2/G\&P) \cdot \Pr(O_n/G\&P\&E_2)
= m_1/(m_1 + m_2) \cdot \Pr(O_n/G\&P\&E_1) + m_2/(m_1 + m_2) \cdot \Pr(O_n/G\&P\&E_2)$$
Now see

Now, suppose that tampering occurs in E_2 so that there are m_3 pseudomelanics in E_2 . We can be define a relative to the E_2 . We can write $\Pr(O_n/P)$ as a weighted average of the probabilities relative to the

$$\Pr(O_n/P) = \Pr(E_1/P) \cdot \Pr(O_n/P \& E_1) + \Pr(E_2/P) \cdot \Pr(O_n/P \& E_2).$$

By the argument that Brandon uses to motivate his claims about screening off, we can take $P_{1}(P) = P_{2}(P) = P_{3}(P)$ can take $\Pr(O_n/G\&P\&E_i) = \Pr(O_n/P\&E_i)$ for i = 1, 2. However, $\Pr(E_1/P) = m_1/(m_1 + m_2 + m_3)$ so that $\Pr(E_1/P) \neq m_1/(m_1 + m_2 + m_3)$ so that $\Pr(E_1/P) \neq m_1/(m_1 + m_2 + m_3)$ so that $\Pr(E_1/P) \neq m_1/(m_1 + m_2 + m_3)$ so that $\Pr(E_1/P) \neq m_1/(m_1 + m_2 + m_3)$ so that $\Pr(E_1/P) \neq m_1/(m_1 + m_2 + m_3)$ so that $\Pr(E_1/P) \neq m_1/(m_1 + m_2 + m_3)$ so that $m_1/(m_1 + m_2 + m_3)$ and $\Pr(E_1/P) = (m_2 + m_3)/(m_1 + m_2 + m_3)$, so that $\Pr(E_1/P) \neq \Pr(E_1/C_2, P)$. $\Pr(E_i/G\&P)$. Thus, $\Pr(O_n/G\&P) \neq \Pr(O_n/P)$, and the claim about screening off fails. Digitized by Arya Samaj Foundation Chennal and eGangotri that Brandon intends to relativize to a fixed environment. But now he has effectively begged the question against the genic selectionist by deploying the orthodox conception of environment. Genic selectionists will also want to relativize to the environment, but they should resist the orthodox conception of it. On their view, the probability relations derived by Brandon involve an illicit averaging over environments (see fn. 12). Instead, genic selectionists should propose that the probability of an allele's leaving n copies of itself should be understood relative to the total allelic environment, and that the specification of the total environment ensures that there is no screening off of allelic properties by phenotypic properties. The probability of producing n copies of the allele for melanin in a total allelic environment is invariant under conditionalization on phenotype.

Here too the moral of our story is that Dawkinspeak must be undertaken consistently. Mixing orthodox concepts of the environment with ideas about genic selection is a recipe for trouble, but we have tried to show how the genic approach can be thoroughly articulated so as to meet major objections. But what is the point of doing so? We shall close with a survey of some advantages and potential

drawbacks.

III. GENES AND GENERALITY

Relatively little fossicking is needed to uncover an extended defense of the view that gene selectionism offers a more general and unified picture of selective processes than can be had from its alternatives. Phenomena anomalous for the orthodox story of evolution by individual selection fall naturally into place from Dawkins' viewpoint. He offers a revision of the "central theorem" of Darwinism. Instead of expecting individuals to act in their best interests, we should expect an animal's behavior "to maximize the survival of genes 'for' that behavior, whether or not those genes happen to be in the body of that particular animal performing it" (EP 223).

The cases that Dawkins uses to illustrate the superiority of his own approach are a somewhat motley collection. They seem to fall into two general categories. First are outlaw and quasi-outlaw examples. Here there is competition among genes which cannot be translated into talk of vehicle fitness because the competition is among co-

Notice that, if environments are lumped in this way, then it will only be under fortuitous circumstances that the tampering makes the probabilistic relations come out as Brandon claims. Pseudomelanics would have to be added in both environments so that the weights remain exactly the same.

builders of a single vehicle. The second group comprises "extended phenotype" cases, instances in which a gene (or combination of genes) has selectively relevant phenotypic consequences which are not traits of the vehicle that it has helped build. Again the replication potential of the gene cannot be translated into talk of the adaptedness of its vehicle.

We shall begin with outlaws and quasi outlaws. From the perspective of the orthodox story of individual selection, "replicators at different loci within the same body can be expected to 'cooperate'." The allele surviving at any given locus tends to be one best (subject to all the constraints) for the whole genome. By and large this is a reasonable assumption. Whereas individual outlaw organisms are perfectly possible in groups and subvert the chances for groups to act as vehicles, outlaw genes seem problematic. Replication of any gene in the genome requires the organism to survive and reproduce, so genes share a substantial common interest. This is true of asexual reproduction, and, granting the fairness of meiosis, of sexual reproduction too.

But there is the rub. Outlaw genes are genes which subvert meiosis to give them a better than even chance of making it to the gamete, typically by sabotaging their corresponding allele (EP 136). Such genes are segregation distorters or meiotic drive genes. Usually, they are enemies not only of their alleles but of other parts of the genome, because they reduce the individual fitness of the organism they inhabit. Segregation distorters thrive, when they do, because they exercise their phenotypic power to beat the meiotic lottery. Selection for such genes cannot be selection for traits that make organisms more likely to survive and reproduce. They provide uncontroversial cases of selective processes in which the individualistic story cannot be told.

There are also related examples. Altruistic genes can be outlaw-like, discriminating against their genome mates in favor of the inhabitants of other vehicles, vehicles that contain copies of themselves. Start with a hypothetical case, the so-called "green beard" effect. Consider a gene Q with two phenotypic effects. Q causes its vehicle to grow a green beard and to behave altruistically toward green-bearded conspecifics. Q's replication prospects thus improve, but the particular vehicle that Q helped build does not have its prospects for survival and reproduction enhanced. Is Q an outlaw not just with respect to the vehicle but with respect to the vehicle builders? Will there be selection for alleles that suppress Q's effect? How the selection process goes will depend on the probability that Q's cobuilders

Digitized by Arya Samaj Foundation Chennai and eGangotri are beneficiaries as well. If Q is reliably associated with other gene kinds, those kinds will reap a net benefit from Q's outlawry.

So altruistic genes are sometimes outlaws. Whether coalitions of other genes act to suppress them depends on the degree to which they benefit only themselves. Let us now move from a hypothetical

example to the parade case.

Classical fitness, an organism's propensity to leave descendants in the next generation, seems a relatively straightforward notion. Once it was recognized that Darwinian processes do not necessarily favor organisms with high classical fitness, because classical fitness ignores indirect effects of costs and benefits to relatives, a variety of alternative measures entered the literature. The simplest of these would be to add to the classical fitness of an organism contributions from the classical fitness of relatives (weighted in each case by the coefficient of relatedness). Although accounting of this sort is prevalent, Dawkins (rightly) regards it as just wrong, for it involves double bookkeeping and, in consequence, there is no guarantee that populations will move to local maxima of the defined quantity. This measure and measures akin to it, however, are prompted by Hamilton's rigorous development of the theory of inclusive fitness (in which it is shown that populations will tend toward local maxima of inclusive fitness). 13 In the misunderstanding and misformulation of Hamilton's ideas, Dawkins sees an important moral.

Hamilton, he suggests, appreciated the gene selectionist insight that natural selection will favor "organs and behavior that cause the individual's genes to be passed on, whether or not the individual is an ancestor" (EP 185). But Hamilton's own complex (and much misunderstood) notion of inclusive fitness was, for all its theoretical importance, a dodge, a "brilliant last-ditch rescue attempt to save the individual organism as the level at which we think about natural selection" (EP 187). More concretely, Dawkins is urging two claims: first, that the uses of the concept of inclusive fitness in practice are difficult, so that scientists often make mistakes; second, that such uses are conceptually misleading. The first point is defended by identifying examples from the literature in which good researchers have made errors, errors which become obvious once we adopt the gene selectionist perspective. Moreover, even when the inclusive fit-

¹³ For Hamilton's original demonstration, see "The Genetical Evolution of Social Behavior I," in G.C. Williams, ed., *Group Selection* (Chicago: Aldine, 1971), pp. 23–43. For a brief presentation of Hamilton's ideas, see Kitcher, *op.cit.*, pp. 77–87; and for penetrating diagnoses of misunderstandings, see A. Grafen, "How Not to Measure Inclusive Fitness," *Nature*, CCXCVIII (1982): 425/6; and R. Michod, "The Theory of Kin Selection," in Brandon and Burian, *op.cit.*, pp. 203–237.

ness calculations make the right predictions, they often seem to mystify the selective process involved (thus buttressing Dawkins's second thesis). Even those who are not convinced of the virtues of gene selectionism should admit that it is very hard to see the reproductive output of an organism's relatives as a property of that organism.

Let us now turn to the other family of examples, the "extended phenotype" cases. Dawkins gives three sorts of "extended" phenotypic effects: effects of genes—indeed key weapons in the competitive struggle to replicate—which are not traits of the vehicle the genes inhabit. The examples are of artifacts, of parasitic effects on host bodies and behaviors, and of "manipulation" (the subversion of an organism's normal patterns of behavior by the genes of another organism via the manipulated organism's nervous system).

Among many vivid, even haunting, examples of parasitic behavior, Dawkins describes cases in which parasites synthesize special hormones with the consequence that their hosts take on phenotypic traits that decrease their own prospects for reproduction but enhance those of the parasites (see, for a striking instance, EP 215). There are equally forceful cases of manipulation: cuckoo fledglings subverting their host's parental program, parasitic queens taking over a hive and having its members work for her. Dawkins suggests that the traits in question should be viewed as adaptations—properties for which selection has occurred—even though they cannot be seen as adaptations of the individuals whose reproductive success they promote, for those individuals do not possess the relevant traits. Instead, we are to think in terms of selectively advantageous characteristics of alleles which orchestrate the behavior of several different vehicles, some of which do not include them.

At this point there is an obvious objection. Can we not understand the selective processes that are at work by focusing not on the traits that are external to the vehicle that carries the genes, but on the behavior that the vehicle performs which brings those traits about? Consider a spider's web. Dawkins wants to talk of a gene for a web. A web, of course, is not a characteristic of a spider. Apparently, however, we could talk of a gene for web building. Web building is a trait of spiders, and, if we choose to redescribe the phenomena in these terms, the extended phenotype is brought closer to home. We now have a trait of the vehicle in which the genes reside, and we can tell an orthodox story about natural selection for this trait.

It would be tempting to reply to this objection by stressing that the selective force acts through the artifact. The causal chain from the gene to the web is complex and indirect; the behavior is only a part of

Digitized by Arya Samaj Foundation Chennai and eGangotri

it. Only one element of the chain is distinguished, the endpoint, the web itself, and that is because, independently of what has gone on earlier, provided that the web is in place, the enhancement of the replication chances of the underlying allele will ensue. But this reply is exactly parallel to the Mayr-Gould-Brandon argument discussed in the last section, and it should be rejected for exactly parallel reasons.

The correct response, we believe, is to take Dawkins at his word when he insists on the possibility of a number of different ways of looking at the same selective processes. Dawkins's two main treatments of natural selection, SG and EP, offer distinct versions of the thesis of genic selectionism. In the earlier discussion (and occasionally in the later) the thesis is that, for any selection process, there is a uniquely correct representation of that process, a representation which captures the causal structure of the process, and this representation attributes causal efficacy to genic properties. In EP, especially in chapters 1 and 13, Dawkins proposes a weaker version of the thesis, to the effect that there are often alternative, equally adequate representations of selection processes and that, for any selection process, there is a maximally adequate representation which attributes causal efficacy to genic properties. We shall call the strong (early) version monist genic selectionism and the weak (later) version pluralist genic selectionism. We believe that the monist version is faulty but that the pluralist thesis is defensible.

In presenting the "extended phenotype" cases, Dawkins is offering an alternative representation of processes that individualists can redescribe in their own preferred terms by adopting the strategy illustrated in our discussion of spider webs. Instead of talking of genes for webs and their selective advantages, it is possible to discuss the case in terms of the benefits that accrue to spiders who have a disposition to engage in web building. There is no privileged way to segment the causal chain and isolate the (really) real causal story. As we noted two paragraphs back, the analog of the Mayr-Gould-Brandon argument for the priority of those properties which are most directly connected with survival and reproduction—here the webs themselves-is fallacious. Equally, it is fallacious to insist that the causal story must be told by focusing on traits of individuals which contribute to the reproductive success of those individuals. We are left with the general thesis of pluralism: there are alternative, maximally adequate representations of the causal structure of the selection process. Add to this Dawkins's claim that one can always find a way to achieve a representation in terms of the causal efficacy of genic properties, and we have pluralist genic selectionism.

Pluralism of the kind we espouse has affinities with some traditional views in the philosophy of science. Specifically, our approach is instrumentalist, not of course in denying the existence of entities like genes, but in opposing the idea that natural selection is a force that acts on some determinate target, such as the genotype or the phenotype. Monists err, we believe, in claiming that selection processes must be described in a particular way, and their error involves them in positing entities, "targets of selection," that do not exist.

Another way to understand our pluralism is to connect it with conventionalist approaches to space-time theories. Just as conventionalists have insisted that there are alternative accounts of the phenomena which meet all our methodological desiderata, so too we maintain that selection processes can usually be treated, equally adequately, from more than one point of view. The virtue of the genic point of view, on the pluralist account, is not that it alone gets the causal structure right but that it is always available.

What is the rival position? Well, it cannot be the thesis that the only adequate representations are those in terms of individual traits which promote the reproductive success of their bearers, because there are instances in which no such representation is available (outlaws) and instances in which the representation is (at best) heuristically misleading (quasi-outlaws, altruism). The sensible rival position is that there is a hierarchy of selection processes: some cases are aptly represented in terms of genic selection, some in terms of individual selection, some in terms of group selection, and some (maybe) in terms of species selection. Hierarchical monism claims that, for any selection process, there is a unique level of the hierarchy such that only representations that depict selection as acting at that level are maximally adequate. (Intuitively, representations that see selection as acting at other levels get the causal structure wrong.) Hierarchical monism differs from pluralist genic selectionism in an interesting way: whereas the pluralist insists that, for any process, there are many adequate representations, one of which will always be a genic representation, the hierarchical monist maintains that for each process there is just one kind of adequate representation, but that processes are diverse in the kinds of representation they demand. 14

In defending pluralism, we are very close to the views expressed by Maynard Smith in "How To Model Evolution." Indeed, we would like to think that Maynard Smith's article and another in a number of Smith's article and the present essay complement one another in a number of respects. respects. In particular, as Maynard Smith explicitly notes, "recommending a plurality of many larger and the present essay compensation of the rality of models of the same process" contrasts with the view (defended by Gould

Just as the simple orthodoxy of individualism is ambushed by outlaws and their kin, so too hierarchical monism is entangled in spider webs. In the "extended phenotype" cases, Dawkins shows that there are genic representations of selection processes which can be no more adequately illuminated from alternative perspectives. Since we believe that there is no compelling reason to deny the legitimacy of the individualist redescription in terms of web-building behavior (or dispositions to such behavior), we conclude that Dawkins should be taken at face value: just as we can adopt different perspectives on a Necker cube, so too we can look at the workings of selection in different ways (EP ch. 1).

In previous sections, we have tried to show how genic representations are available in cases that have previously been viewed as troublesome. To complete the defense of genic selectionism, we would need to extend our survey of problematic examples. But the general strategy should be evident. Faced with processes that others see in terms of group selection or species selection, genic selectionists will first try to achieve an individualist representation and then apply the ideas we have developed from Dawkins to make the translation to genic terms.

Pluralist genic selectionists recommend that practicing biologists take advantage of the full range of strategies for representing the workings of selection. The chief merit of Dawkinspeak is its generality. Whereas the individualist perspective may sometimes break down, the gene's eye view is apparently always available. Moreover, as illustrated by the treatment of inclusive fitness, adopting it may sometimes help us to avoid errors and confusions. Thinking of selection in terms of the devices, sometimes highly indirect, through which genes lever themselves into future generations may also suggest new approaches to familiar problems.

But are there drawbacks? Yes. The principal purpose of the early sections of this paper was to extend some of the ideas of genic selectionism to respond to concerns that are deep and important. Without an adequate rethinking of the concepts of population and of environment, genic representations will fail to capture processes that involve genic interactions or epigenetic constraints. Genic selectionism can easily slide into naive adaptationism as one comes to

and by Sober) of "emphasizing a plurality of processes." Gould's views are clearly expressed in "Is A New and General Theory of Evolution Emerging?" *Paleobiology*, vi (1980): 119–130; and Sober's ideas are presented in NS ch. 9.

credit the individual alleles with powers that enable them to operate independently of one another. The move from the "genes for P" locution to the claim that selection can fashion P independently of other traits of the organism is perennially tempting. 15 But, in our version, genic representations must be constructed in full recognition of the possibilities for constraints in gene-environment coevolution. The dangers of genic selectionism, illustrated in some of Dawkins's own writings, are that the commitment to the complexity of the allelic environment is forgotten in practice. In defending the genic approach against important objections, we have been trying to make this commitment explicit, and thus to exhibit both the potential and the demands of correct Dawkinspeak. The return of the gene should not mean the exile of the organism. 16

KIM STERELNY

Victoria University/Wellington, New Zealand

PHILIP KITCHER

University of California/San Diego

op.cit.

16
As, we believe, Dawkins himself appreciates. See the last chapter of EP, especially his reaction to the claim that "Richard Dawkins has rediscovered the organism" (951)

¹⁵ At least one of us believes that the claims of the present paper are perfectly compatible with the critique of adaptationism developed in Gould and Lewontin, "The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme," in Sober, ed., Conceptual Problems in Evolutionary ism, see Kitcher, Vaulting Ambition, ch. 7; and "Why Not The Best?" in Dupre, op.cit.

COPING WITH CONFLICT

his paper discusses some ways of dealing with the existence of conflict. The topic here is hardly new; Plato considers it in the *Republic*. Many philosophers have written about it, and all of us have studied it. Conflict remains a fact of life, but that cannot be blamed on us. Our business is not to solve social problems but only to report and to understand them. Still, even so, on the matter of conflict, we have done rather poorly. We have focused on one problem only, a very important problem. We have ignored a second problem that is just as important.

I will be speaking mainly of the conflicts of separate people. There is an ambiguity here, for people's conflicts are of two sorts. Some of them are with other people and some are conflicts we have with ourselves. Some are interpersonal and some are *intra*personal—I will call them *social* and *psychological*. What follows is mainly about social conflicts, but the ways of coping with these two kinds of conflict are similar.

What do I mean by "conflict"? A person has a conflict with himself where he wants two things (or propositions) he knows can't both be had (or made true). Such a person wants a and also b and believes that if a then not-b, and thus if b then not-a. In a conflict of the social sort, this is parcelled out among several people. In such a case, some people want a and others want b and each of them thinks that both can't be had.

On this idea of conflict, a conflict need not be an opposition of values: not every conflict is a moral conflict. If the people at odds see what they want as morally proper or right, if their values demand that they get it, they are engaged in a moral conflict, but many conflicts are not like this. Where you want us to move to Utah and I want us to move to Maine, we are in a conflict that involves no morality. Most of the recent studies of conflict are about moral conflicts only. These are special cases of my more general concept.

Where now is the problem of conflict? The usual answer is that a conflict makes for a standstill, that it inhibits action. Not every conflict is like this either. The forces released push against each other, but they don't always cancel out. It can happen that people in a conflict don't even know they are in it, or don't care enough to make an effort, or are too weak to block the others. It may be too that none of the parties knows how to bring about what he wants (you want fair weather, I want rain). Still, the point is important: conflicts often paralyze action. Something or other has to be done and people

0022-362X/88/8507/0362\$01.40

© 1988 The Journal of Philosophy, Inc.

stand in each other's way. I will call this the *Problem of Action*. Two ways of coping with it often are noted. The first I will call the *Method of Inquiry*.

In a conflict we disagree, but of course that need not continue. We might yet come to agree; that would then end the conflict. The method of inquiry is to try for just this. Here is a quote from John Dewey, who gave it the name I am using. Suppose that there are

. . . conflicting desires and alternative apparent goods. What is needed is to find the right course of action, the right good. Hence, inquiry is exacted: observation of the detailed makeup of the situation; analysis into its diverse factors; clarification of what is obscure; discounting the more insistent and vivid traits; tracing the consequences of the various modes of action that suggest themselves . . . (133).

This has the ring of common sense, and it is sometimes the natural course. Where the desires in conflict rest on opposing beliefs or judgments, focusing on the beliefs involved may be the thing to do. Comparing the evidence relevant to them could dissolve the belief opposition and so resolve the conflict of desires. Suppose that the conflict is being fought over the marketing of some drug. Those who want it kept off the market think that the drug is dangerous. Those who want it made available think that the drug is safe. Let them get together and jointly study the evidence, design new experiments, analyze the data, etc. If they follow the same logic in this, they may come to agree.²

But often a conflict does not derive from any difference in people's beliefs. Should the U.S. support the Contras? Should testing for AIDS be required by law? Those who say yes and those who say no may hold the same beliefs on the facts. Still, they remain in conflict, and their conflict may persist however much they yet learned. Here the method of inquiry holds very little promise. Or think of capital punishment. Where people disagree on public hanging because they differ on its deterrent effect, studying the evidence may settle their conflict. Where they don't agree on hanging because they differ on the morality of revenge, no amount of inquiry will provide for agreement.

It may be said in defense of the method that it here isn't given a chance. For an inquiry to do its job, a desire-neutral setting must be found for it. That is, the parties seeking agreement must start out on

¹ Reconstruction in Philosophy (New York: Mentor, 1950).

This speaks of social conflicts only. Making it bear on the psychological sort calls for taking a person to have opposing beliefs on his own. That makes little sense to me; per contra, see Isaac Levi, Hard Choices (New York: Cambridge, 1986).

common ground, not in separate ball parks, as they do in the revenge case above. They must forget or ignore not only the desires that make for the conflict between them but all desires and interests that might keep them apart. This has been urged as the proper course where the issue is that of basic social institutions—think of John Rawls's idea of asking how we would vote behind a veil of ignorance. But no one thinks it proper to shelve all opposing interests whatever the issue. Where we differ on whether to require all people to be tested for AIDS, should our inquiry start by excluding the likely harassment of those found infected, which (let us say) matters to me but does not matter to you? If we came to agree on the tests by ignoring this deeper opposition of ours, what force would our agreement then have? Why should I think myself bound by it when I returned to my senses?

Still, the method may work where the conflict rests on some difference of opinion. Suppose that we disagree on the testing because you think it more likely than I do that AIDS might become epidemic; in that case, collecting more evidence could bring us into agreement. But a problem remains even here. We might agree in the fullness of time, but that may be more time than we have. Action may have to be taken while we still are in conflict on the matter. We need here a second method of coping I will call *Arbitration*.

This is any fixed procedure for selecting an outcome. There are very many such. Think of two salesmen competing for a sale. They lower their prices or sweeten the deal; the customer then decides. Or take a trial in court. The jury brings in a verdict, and that determines who wins. In an issue of political leadership, the citizens go to the polls. Where all else fails, people slug it out—this is the oldest arbitration procedure. Sometimes, as in competition or in court, a third party enters and settles it. Sometimes, as in physical combat, no one else is called in.

In an inner conflict, we may toss a coin; this too is an arbitration. Notice that often the conflict remains though the course to be taken is set. Having tossed the coin, I still want to go to Japan and also want to go to Spain, but it fell heads and so Spain it is. So also on the larger scene. Having heard how the vote came out, all the voters accepted the outcome, but some of us still wanted Mondale. (We put this in the subjunctive: we wished that it had been he.) Where an issue of action is settled, the parties may remain at odds, or remain at odds with themselves.

Sometimes we resort to a hypothetical arbitration. Suppose that there is a pie, and that both of us want it all. Strong words are exchanged; neither one of us budges. We might now imagine a mag-

istrate, an informed, impartial observer. We ask ourselves what he would say and agree to a compromise: we split the pie. Each of us still wants it all—we wish we *could* have it all—but we each settle for half. (In a compromise, all parties back off, no one gets what he wants; we will ignore such cases below and keep to what happens where one party wins.)

Not every arbitration is acceptable to us. A dictator imposes solutions, but that is not how we want it done. All too often, the dictator's friends wind up with most of the pie. The verdict of force may have to be followed but it cannot be justified, and we want an arbitration to be just or fair. Here we meet the troubling question of the nature of fairness. There are, indeed, many different questions: What is fair competition? What is a fair trial? Perhaps even, what is fair combat? Though strictly, we look for fairness only in the settling of social conflicts; in an inner, psychological conflict, the issue of fairness is never raised. We think here more in Platonic terms, looking for wisdom rather than fairness. Tossing a coin won't do at all in any serious inner conflict. Hamlet would have rejected the idea of letting heads or tails settle things.

Arbitration works; it solves the problem of action. (That is, it works where we let it work.) But it leaves a second problem I have not yet mentioned, a problem philosophers seldom consider. Again, a conflict may make for a standstill, and arbitration provides against that. One side wins and the other side loses, and so the course is clear. But there is something else besides. One party goes off whistling; the other goes off grumbling. Arbitration leaves people unhappy, and this both in social and psychological conflicts.

Take a difficult tenure issue. The department is split and both sides feel strongly. They argue, but the division holds. When the vote is taken, the majority is against. Arbitration has worked: the divided department has reached a decision. But of course the conflict remains, and now it may take on some bitterness. There will be repercussions. The trouble is far from over.

Or take a psychological conflict. Suppose that Hamlet did toss a coin. Suppose that it came out for *not* killing Claudius and that Hamlet abided by this. Would he then have been free and clear? His conflict would have remained: he still would have wanted to kill his uncle and also to keep the love of his mother and known that each excluded the other. Could the play have ended right there?

³ This may depend on the issue and on who the dictators are—call them, as Plato did, guardians. See Robert Dahl, Controlling Nuclear Weapons: Democracy Versus Guardianship (Syracuse: University Press, 1985).

Hamlet of course was neurotic. Shouldn't a person who tosses a coin allow the fall to settle his mind? Also, in the tenure case, having decided to take a vote, should we not accept the vote's outcome? Yes we should, but often we don't, and neurosis has nothing to do with it. The point is again that an arbitration does not always put an end to the conflict. There is then dissatisfaction and sometimes also resentment. The problem is what to do about it.

One problem of conflict is that of action. Another is that of making peace. Or we might put it differently, that we see here an ambiguity in the *ending* of conflicts. To *end* a conflict might just mean to get one party to stand aside, to bring it about that this party loses. Or it might mean to bury the hatchet, to dispel the bad feelings the loss left behind. We have, on one hand, the *settling* of conflicts; on the other, their being *retired*. The methods of inquiry and arbitration are different ways of settling conflicts. They leave the conflicts unretired, and in that sense, still unresolved. The parties then remain at odds—strictly, their conflict remains. They must still try to make peace.

It may be asked, why bother? Suppose there has been some arbitration. The question of what to do is settled: we are finally on the move. Suppose that the process was fair. Why not just get on with it now? Why worry about sore losers? Why the winners should worry is clear. The losers might make it hard for them. At any rate, they may be uneasy with the resentment of those they defeated. But these others too may want the issue put behind them; they may dislike being angry. Both the winners and the losers may want to make peace, and both for good reasons.

Most philosophers ignore making peace. This because, when they write about conflict, it is justice alone that concerns them. Where an issue was settled unfairly, justice has yet to be done. Perhaps some compensation is called for, perhaps retaliation. Perhaps the wrong can never be righted and justice is out of reach. But where justice has been done, the job is finished, the case is closed—this is the usual view of it.

Justice has nothing to do with our problem, no one here having been wronged. The issue was settled by arbitration, and the procedure used was fair. The problem of making peace remains, for having lost can itself be hard. It can leave the losers bitter and turn them against all the others. This would make for what Hobbes called war, "[which] consisteth not in actual fighting, but in the known disposition thereto." Hobbes and some other contract theorists noted a

⁴ Leviathan, part I, ch. 13.

problem much like ours, but they defined it more narrowly. By "actual fighting" they meant knives and clubs, and so they held that what was needed was to establish a strong police. Where we hope for more from peace than the avoidance of bloodshed, we must do more. We cannot just hire a guard. We are likely to try something else I will call *Persuasion*.

This way of coping with conflict is common but is hard to explain. It aims at a sort of agreement, but not as the method of inquiry does. In that, the project was settling the conflict by smoothing out the differences. Persuasion starts off from the premise that such agreement is not to be had. But if that is not its purpose, what is the point of persuasion? It may be well for us to step back to get more perspective here.

Some months ago, the *New York Times* printed an article by David Callahan⁵ against the legal endorsement of contracts arranging for surrogate motherhood. Callahan held that endorsing these contracts would create "a cadre of women whose prime virtue is what we now take to be a deep vice—the bearing of a child one does not want and is prepared not to love" (*ibid.*). A few weeks later, the *Times* printed a letter responding to Callahan's article.⁶ The letter argued that a surrogate mother "is not discarding a child; she is bringing the gift of a most wanted child to its father and adoptive mother, to be loved [by them]" (*ibid.*). Here were two tries at persuasion. What were the authors doing?

Neither was offering any information he thought the readers might not have. Nor need it be that either was trying to change what the readers wanted, to shake any reader loose from how he wanted the surrogate issue to go. They were reshaping that issue, or rather, trying to replace it with another, a different but closely related one. They were trying to change the agenda.

An issue is a set of possibilities or propositions, some one of which must be made actual or true, and it hasn't yet been decided which. The basic surrogate mother issue is this two-membered set: contracts for surrogate mothering will be allowed, no such contracts will be allowed. In his article, Callahan tried to get us to focus on this issue instead: we will create a group of women who will not love their children, we will not do this. The letter writer tried to make it this: we will provide for people getting the children they want and wouldn't otherwise have, we will not provide for it.

These issues are of course related. Their different first options

⁵ "Surrogate Motherhood: A Bad Idea" (Jan. 20, 1987). Feb. 13, 1987.

Digitized by Arya Samaj Foundation Chennai and eGangotri

pick out the same course, and so do their second options. Under present circumstances, allowing the contracts involved is the same as creating a group of women who will not love their children, and also the same as providing for people getting the children they want and would not otherwise have. And *not* allowing these contracts is the same as *not* creating this group and also *not* serving these people. The issues differ in this, that each lays out the prospects differently, that is, under different propositions.

Why is this difference important? Why should people bother to promote one issue over the other, to turn our minds to just these propositions, to try to persuade us to think in their terms? Because they believe that our thinking that way would connect with desires we had all along, desires we share with these people. Changing the agenda is changing the subject, and on the subject they are bringing up for us, these people know we agree with them. By getting us to accept their agenda, they hope to get us to join forces with them.

Persuasion rests on a philosophical insight. It rests on a point about the semantics of wanting, that what a person wants are not events or situations in the physical raw, but these events or situations under some descriptions of them, that what we want in every case is that a certain proposition be true. A new agenda has new propositions, and so connects with different desires, perhaps desires at variance with those that moved us before. In that case, we want something under one description and do not want it under another. We may even want a proposition true and want a coreportive proposition false, and this where we know that the propositions are coreportive (that they report the same).

Let me refer to the way that a person describes things to himself as his understanding of them. Persuasion is the attempt to change a person's understanding of something, to get him to see it in some way that prompts him to act as he would not have done. Callahan hoped to get us to think of surrogate motherhood as loveless childbearing. He knew that we were opposed to the latter. If we came to see it as he did, we would all stand together. We would vote (or whatever) against it. The second writer assumed that we wanted couples who wanted children to have them. If he could get us to think of surrogacy as obliging these people, we would then side with him.

Persuasion here is a predecision strategy; it gets us to turn away from one issue and to attend to another instead. So let us leave this example. We want to see what happens later, after the matter is settled, to see how persuasion makes peace. In the familiar scenario, the man brings home his new wife and says, "Think of it this way,

Mamma. You're not losing a son. You're gaining a daughter." Persuading here is reconciling. A decision has been made and acted upon. We persuade the party it disappoints by putting the outcome so that they can accept it.

Sometimes it is we who are disappointed. How does a person persuade himself? Suppose that some vote went against my side. I may now remind myself that I argued the best I could, that the vote was fair, and that the person who was elected has the support of the group as a whole. I still don't like him, but I then see him as the group choice and so am reconciled to him. That is, I can live with the outcome—I stop chafing at it.⁷

Here is the same with an outside persuader. My side has lost but I don't go along. Instead, I make what trouble I can. My friend (or someone from the other side) now takes me to task for this. He calls my behavior pigheaded. What I am opposing is the popular will. Should the result of a vote be ignored? Why should I count for more than the others? What, after all, is democracy about? These questions change how I look at the outcome and I come to accept it.

Being persuaded implies accepting what before we did not. What is accepting a situation? Take any situation that a person thinks holds. He accepts it where he understands it as the sort of thing that he wants. Unlike wanting, which focuses on propositions, accepting has events and situations as its objects. A person may accept a situation though he diswants (wants false) some proposition that he thinks expresses it. In such a case, he understands the situation under a different description and he wants the proposition corresponding to that second description. In the example above, I wanted a certain person to lose (I diswanted him to win) but came to see his election as a case of majority rule. I want the majority to get its way, and so I accepted his winning.

A different case from literature: Rubashov in *Darkness at Noon*. The accusation against him is false. Rubashov insists on his innocence. The public examiner tells him that the issue is closed—the verdict will be *guilty*. The truth does not matter, he says, for this verdict will serve the cause. He argues that a confession would do even better. It would shock people into a sense of the danger the country is in. Rubashov comes to see confessing as a service to the Party. He accepts his being convicted and agrees to sign.

Or take a self-persuasion again. An army officer is assigned to a project he thinks is excessively brutal. He does not want to go

 $^{^7}$ Who was the I that needed persuading and who the I that persuaded him? How were the two (the one!) related? These matters are not for this paper.

Digitized by Arya Samaj Foundation Chennai and eGangotri

through with it, but that issue is settled: he *must*. He reminds himself now that he is a junior officer and that it is not his job to decide but to follow orders. He has been given instructions, and, for a soldier, that should suffice. Arguing thus, he persuades himself. He sets aside his objection to the project and accepts his involvement in it.

But should he have changed his mind in this way? Ought a person to see brutality as just following orders? The man has in fact come to see it so, which means that it isn't a mere excuse. Nor is it self-deception—he will indeed be following orders. The question is, has he misled himself? Has he twisted his head around from where it should have been facing?

What about Rubashov now? Should he have seen his confession as a service? Again, he was not deceived. He did in fact serve the cause by lying. Still, was he being a martyr here or was he just being had? And in my voting example: should I have seen the voting outcome as an expression of the popular will? The answers we get in these cases may differ, but the question is the same each time. When is a description of a situation proper? We are assuming it to be correct, but when is the description appropriate? When is a person representing things rightly, and when is he misrepresenting them?

Persuasion can sometimes mislead, and this both where it is self-addressed and where it is addressed to another. The important question this raises is, when does it and when doesn't it? This corresponds to some questions the other methods raise. The question regarding inquiry is: when is an inquiry conclusive? When have we collected information that should lead us to choose (or to want) a certain option? This is the problem of decision theory. The question of arbitration is: when is it fair? That is the subject of social-choice theory. The question of when a persuasion misleads is not the concern of any field of study. It seems never to have been studied at all.

The question boils down to this: when are the facts being misrepresented? When are we not understanding things rightly, and when might we see them as we do? This goes beyond the issue of conflict. It is a general philosophical question, a question like that of warranted belief, which has an extensive literature. The logic of understanding has none, and this has to do with the common conflation of understanding and belief, with the idea that an understanding of something is just what the agent believes about it. The first step is to distinguish these two, and this was taken by Aristotle, who referred to understanding as "judgment." But Aristotle begged our present

⁸ There are only remarks like Dewey's about "discounting the more insistent and vivid traits"; see the quote above.

question by supposing that judgment was always sound—a person of judgment understood things rightly. So we will get no help from him.9

Let me shelve this difficult question and return to persuasion as a way of making peace. That is, let me say that it may not work. Those whom we want to see persuaded may resist our persuasion. The arbitration was fair, but those who lost may think that it wasn't and so refuse to listen. Or the issue was a moral one for them, and to yield to persuasion now strikes them as yielding to evil. Or perhaps they simply don't find the understanding we offer them any better than their own. It may be too that we brought them over, but that the persuasion did not stick, that they initially accepted the outcome as we described it but no longer do, having thought it over and come to take a different view of it.

Or it may even be this, that the persuasion did stick but that the desire then set aside is now fighting back. The agent accepts the outcome arranged. The way he now sees it, he can't object to it; it does not look objectionable to him. This the persuasion achieved. Yet he cannot stomach it either—his gorge rebels against his mind. Some young twerp, fresh out of school, was given a job that should have been his. He no longer thinks this unfair: he was persuaded that it served a good purpose. But he remains (or has become) envious and angry. Social psychologists sometimes use the Nietzschean term 'ressentiment' for this. Psychoanalysts describe it as "the return of the repressed." Here is a species of inner disorder philosophers never consider. 10

How this disorder ought to be studied is of course not clear. But set that aside—my point is only that a persuasion needn't work. We may still face bitterness. More persuasion might now be tried, but the prospects of this are not good. A different approach is in order. We come here to a fourth mode of coping I will call Accommodation.

This takes account of the fact that a conflict is often part of a history. You and I differ in what we now want, and one of us has to lose. But this may not be the last time we will meet. Accommodation starts off from the premise that we can't make peace here today. It looks instead to what might be done somewhere down the road. Better: the winner looks down the road. Unlike the case with per-

For more on the soundness of understandings, see my "Under which Descriptions?" in Amartya Sen and Bernard Williams, eds., Utilitarianism and Beyond (New York: Cambridge, 1982), pp. 251–260. See also my "Rationality: A Third Dimension," *Economics and Philosophy*, III (Spring 1987): 49–66.

Never, with some exceptions; see Max Scheler, *Ressentiment* (New York: Schocker, 1970)

suasion, which losers often address to themselves, accommodating the losers is only open to the people who won.

Suppose that a department is in a conflict over a certain candidate for appointment. The majority is opposed to appointing him; they think him not good enough. The minority is strongly for him because he is black and the department has no blacks. The vote goes against the candidate, and now there is anger and alienation.

The issue was about this person, but the minority's interests went deeper, that is, the minority wanted this person because of what getting him would have meant. The heart of the matter was race. These people had wanted to take a first step toward a more balanced department, and they lost on this deeper intention as well as on the surface issue. The others could move to accommodate here. If they did not oppose integration but just this particular candidate, they might now find an occasion to integrate with someone else they take to be qualified. That is, they might design a new issue on which there would be no conflict and on which the deeper intention of the minority on the previous issue would win—this because they (the initial majority) would now have joined the others.

Is accommodating a kind of payoff, a making it up to the losers for their loss by giving them something else? No, for what the losers will gain is closely related to what they had wanted. They will gain their basic objective, what had led them to want the outcome they had been pressing for. Accommodation goes under the surface and attends to the losers' deeper desires. In finding a way to gratify these, it undoes part of the loss.

A problem has to be faced here too. Let me change the appointment story. Suppose that the candidate being considered was a black scholar who was first rate, the majority pressing for his appointment, the minority opposing it on the grounds of his color. Having appointed this candidate, should the majority now find an occasion on which they could side with the bigots? To accommodate a person or group, the others must do what they would not do if making peace didn't call for it. This means that they must pay a price, and the price is sometimes too high.

When is the price acceptable? There is a clear and simple analysis and also a messier one. The clear analysis takes into account what the accommodation leads the agents to forgo. In the first of the integration stories, the majority shared the minority's desire to move toward integration, so their later voting for that did not itself impose any

I am calling a desire for x deeper than a desire for y where the agent would not want y if he did not want x and the converse is false.

cost. They had to take the time, however, to look for a suitable black appointee, which they would not have normally done, having other priorities. So the cost was this: they had to pass up some other pursuits. In the second scenario, voting the next time with the bigots would not connect with any shared desires. It would mean voting against the next black, whatever his merits were, and that might mean passing up a good colleague. In both scenarios, the question is whether the majority prefers making peace and paying the price to letting things be. If the answer is *yes*, the price is right; if it is *no*, the price is too high.

The messier approach to the costing problem focuses not on the effects of peace making but on the way we would bring peace about. Do we (the agents) really want to side with these or those others? We think of ourselves, in part, in terms of our various commitments and oppositions, and their coherence or hang-together is what we call our *integrity*. Where we agree to side with bigots, is this inner coherence diminished? We would be stepping out of character, but would we be losing integrity? That too would be a price. The question raised here is different: do we prefer establishing peace at the cost of integrity to not having peace?

This question does not help us at all, for we have no handle on it. What is our inner coherence worth? What would it cost to join the others—to do it just this once? We are in murky waters here, the kind that philosophers normally shun. We prefer swimming where we can see bottom. What is the *test* of this sort of coherence? What indeed are the states or conditions the coherence of which is at stake? That is, what sorts of inner states are our commitments and oppositions? Perhaps these questions should be asked first. But however we answer these questions, integrity must come out as something we value, and also as something that can be lost. So the costing of accommodation must take proper account of it. The upshot is the same as above: the price of peace may be too high.

Conflicts on the psychological level also allow for accommodation. Say that I want to work tonight and also to see a movie; I may decide to work and then to take off tomorrow. Or suppose that I want to be a doctor and also to be a musician; I become a doctor but also a zealous amateur fiddler. Is there a problem of costing on this level? Whether my accommodating myself to me can forfeit my integrity is not clear. All depends here on what we mean by commitment and opposition. (Can a person be opposed to what he currently wants?) Either way, the fact remains that I may have to forgo other plans. Supporting the deeper sources or intentions of a desire I overcame may demand a good deal. It might, for a start, require me to go to

some trouble to learn what they are. It might then also oblige me to

change the way that I live.

Accommodation is not always possible. Suppose you came out the winner in some conflict you had with your neighbor: he had objected to your violin. You would now like to make peace with this person, to get it behind you by accommodating him somehow. One of his deeper intentions had been to secure himself against all disturbance, and this you cannot arrange for him. Another had been to prove to the world that he always gets what he wants, which you have shown to be false. A third was perhaps to oppose your will, and this you can't support him in either. He may have had no deeper intention you can now arrange to support. In such a case, you cannot accommodate, no matter what you do.

There are also cases in which the desire that was defeated had no deeper intention. Consider a painful custody fight. Each of the parents wants the child; the court awards it to the father. He cannot then accommodate the mother, for her wanting to have her own child rests on nothing else that she wants. It is, in that sense, a *basic* desire. In such cases too there is nothing one can do. The person who won can't move to make peace, at least not by any accommodation.

A few words ought still to be said about a sort of conflicts I have not discussed. Some of the most far-reaching conflicts are those between institutional units, between unions and companies, or political parties, or sovereign nations—call them *corporate* conflicts. ¹² Here too we need to consider both the problem of action and that of making peace.

Much of what I have said about personal conflicts carries over directly. Again, the method of inquiry works where the conflict rests on a difference of judgment. Take the current conflict between the U.S. and certain countries in Europe over how to deal with the problem of the hole in the ozone layer. This conflict is based on a difference of opinion regarding the role of fluorocarbons. An inquiry is in order here, assuming we have the time for it.

The arbitration of corporate conflicts is a familiar matter. There are panels for labor disputes and bodies like the U.N. to arbitrate national conflicts. And there has always been war, a form of arbitration too. Accommodation is much more rare. The best known in our

¹² Another neglected sort of conflicts are the asymmetrical ones. In these, an institutional unit is pitted against some person or people, for instance, the local telephone company against just you alone. See James S. Coleman, Power and the Structure of Society (New York: Norton, 1974), and his The Asymmetrical Society (Syracuse: University Press, 1982).

century was the Munich Agreement of 1938, a not very likely model of peace making. (The Germans had failed in their special objectives in the First World War; they were here granted their deeper desire of extending their national borders.)

What about persuasion? Can one speak of a group's being brought to change its understanding of something? Can groups be said to have understandings at all? We talk without uneasiness about what groups want and believe, this because we know that such talk can be boiled down to talk about separate people. A group's beliefs and desires are logical constructs of the beliefs and desires of its members, or (more usually) they are the beliefs and desires of certain leaders of the group. The difference of opinion on the role of fluorocarbons is a difference between the U.S. and France only in that certain officials in these countries differ on the subject. To say that the U.K. wants to stay in the Falklands is just to say that some Britishers want this. (Germany wanted to extend its borders; this means that some Germans wanted this.) Likewise here with group understandings. We can construe them distributively: a group's understanding of a situation is the understanding of some of its members. So we need not be uneasy with the idea of persuading a group.

Our methods of coping carry over from the personal to the corporate sphere, and so do their distinctive problems. Persuasion can be misleading here too, and accommodation can ask too much—clearly, at Munich too much was asked. When are we being misled? When is the price we must pay too high? These questions come up for the corporate we as well as for the personal I.

So we meet the same problems again. We have spoken mostly of two ways of making peace, and we now see that they don't always work and that, where they do, they sometimes raise problems. Still, the point of noting the problems was not to despair of the making of peace. Living on decent terms with our neighbors calls for more than just getting our way. It often calls for helping the others over their not having got theirs. How this might be done is not always clear, but that is what makes it a challenge.

Rutgers University

FREDERIC SCHICK

THE JOURNAL OF PHILOSOPHY

Digitized by Arya Samaj Foundation Chennal and eGangotri

BOOK REVIEWS

The Fragility of Goodness. MARTHA NUSSBAUM. New York: Cambridge University Press, 1986. xviii, 544 p. Cloth \$59.50, paper \$19.95.

Martha Nussbaum discusses three aspects of human life which may be recalcitrant to rational planning: (1) Sometimes the values that rational agents pursue turn out to conflict. (2) Sometimes their plans are frustrated by external events beyond their control. (3) Sometimes their rational plans conflict with their nonrational desires and motives (6f). Nussbaum explores these three problems in selected plays of the three major Attic tragedians and in selected passages of Plato and Aristotle. She claims that Plato revolts from the tragic view by denying any value to the recalcitrant aspects of life, and that Aristotle reasserts their value.

Nussbaum offers to present Plato's argument in Republic ii-ix for the claim that happiness is invulnerable to external conditions. She does not show, however, that Plato even accepts this claim. He asserts and defends only the claim that the just person is in all circumstances happier than the unjust person, not the stronger claim the just person is always happy.

Whereas it is not clear that Republic ii-ix accepts the stronger claim, it is fairly clear that the Socrates of Plato's early dialogues (plausibly identified with the historical Socrates) accepts it; for he affirms that virtue is sufficient for happiness, and that no harm can befall a good person.2 It is quite surprising that Nussbaum never mentions or discusses Socrates's view as such. For she seems to assume (without directly facing the issue) that Platonic metaphysics and epistemology are necessary supports for Plato's belief in the invulnerability of virtue. (If they were not necessary, Aristotle's attacks on Platonic epistemology and metaphysics would not be as effective as Nussbaum takes them to be in undermining Platonic

0022-362X/88/8507/0376\$00.90

© 1988 The Journal of Philosophy, Inc.

¹ Much later (495) she cites Republic 388ab. She probably has in mind 387d11-388a1 (partly translated, with a wrong reference, on p. 200). This passage says that the good person will be more self-sufficient than other people (malista), and will be less prone than other people (hêkista) to lament extravagantly at the death of a relative. Nussbaum's suggestion that in this passage Plato asserts the invulnerability of the good person misses the comparative force of malista and hêkista.

² See Plato, Apology 41cd; Crito 48c8-9; Gorgias 470e9-11, 507c. Nussbaum mentions the first passage (495) and attributes the view to Plato. In fact, her view about the historical Socrates is obscure. She mentions only a late (in her view) dialogue, the Phaedrus, to confirm a claim about the historical Socrates (125; I would think the Phaedrus is very poor evidence for her claim).

moral theory.) The Socratic claim about invulnerability shows that Plato at one time accepted invulnerability without the metaphysical background of the Republic. It would be quite relevant for Nussbaum to explore the reasons that might support the Socratic claim. and to explain why Plato is not satisfied with them in the Republic. Her failure to do this is a serious gap in her main argument.

In the passage of Republic IX discussed by Nussbaum (141), Socrates claims to have completed the first proof of his main thesis, and turns to a second proof (580c9-d1). Nussbaum does not even mention this fact; but, if Plato thinks he has proved his main thesis without reference to the considerations peculiar to the second proof. then he cannot believe that his main thesis needs the second proof. I would like to know why Nussbaum disagrees with Plato; for the first proof is ostensibly the culmination of the argument of Books ii-ix, and in ignoring it she ignores most of the argument of the Republic.

She argues that, in Plato's view, nothing has intrinsic (i.e., noninstrumental) value if its value is need-relative (i.e., its being valuable depends on some imperfection of the being for whom it is valuable) (142-147). (1) She takes Socrates in the Gorgias to persuade Callicles that "no activity has positive intrinsic value merely because it replenishes a lack or answers to a need" (143). But Socrates challenges only the different claim that the good consists wholly in first cultivating and then satisfying appetites.3 (2) She claims to find in Republic ii (358b-d) a distinction between "merely context-relative value" and "value that would be value in any circumstances whatever" (145). But the only contrast Plato draws here is between instrumental and intrinsic value; Nussbaum assumes without defense that Plato here denies any context-relative intrinsic value. (3) In Republic ix on false pleasures, Plato suggests that the pleasure experienced only against a background of felt deficiency or pain is less true than the pleasure enjoyed against no such background (146). But to deny some sort of truth to the pleasures is not (as Nussbaum supposes) to deny all intrinsic value to them. In any case, relativity to felt deficiency does not imply anything about relativity to imperfection.

Some further claims about value rest on a broader argument, which I think is this (256f): (1) Plato is a metaphysical realist; he rejects the

³ Moreover, neither Socrates nor Callicles thinks that, as Nussbaum assumes, Callicles' thesis is taken to have been refuted by the mere introduction of allegedly sordid. sordid or disgusting examples. 495b8-c2 makes it clear that Callicles' thesis is not ver refer the callicles. yet refuted. Nussbaum cites 494e7, but overlooks (460n24) the relevance of its immediate sequel.

Digitized by Arya Samaj Foundation Chennai and eGangotri

Protagorean Homomensura doctrine, and so believes that (e.g.) the wrongness of unprovoked murder is not constituted by, or a causal result of, anyone's or everyone's belief that it is wrong. (2) The only epistemological position that could support a claim to have found such truths is nonperspectival and nonanthropocentric, denying that any human agreement (no matter how plausibly supported) constitutes any sort of evidence for a claim to know such truths. (3) Therefore, Plato does not accord any evidential value to such agreements. (4) Aristotle accepts some human agreements as having some evidential value toward his conclusions. (5) Therefore, Aristotle is not a metaphysical realist about his conclusions.

Nussbaum does not defend, or even clearly identify, all the steps of this argument. She does not argue for (2). Her argument for (3) relies on Republic 504, where she finds "Socrates' cryptic claim... that the earlier story was insufficient because it stopped with a merely human agreement and used as its measure of value a human, therefore an imperfect, being" (139). Socrates, however, stops well short of Nussbaum's claim, and hence short of (3). Glaucon and the others were content with an agreement resting on an incomplete examination of the issues; Socrates replies that such an incomplete examination could not be an adequate measure of the truth of the conclusion, and, in general, nothing incomplete is an adequate measure of anything (504e2-3). The fact that one agreement between human beings was an inadequate basis for closing the inquiry does not imply that no human agreement could be an adequate basis. Certainly Socrates does not say that any human agreement constitutes the truth; hence, the passage supports (1). But, unless he accepts (2), it is possible that he is a metaphysical realist who believes that some human agreements give us good reason to believe we have found the truth. This possibility cannot be ruled out, since the passage does not endorse (2). Hence, the passage does not support (3). It would support (3) if we had some independent reason to attribute (2) to Plato; but we have none. Nussbaum, therefore, gives no good reason for (3).

Similarly, her arguments about Aristotle support (4), but not (5). She would have had a case for (5) if she had given some independent reason for attributing (2) to Aristotle; but she gives none. She does not always seem to distinguish (4) from (5), any more than she distinguishes (1) from (3). She therefore gives no reason for believing either that Plato accepts a radically nonperspectival epistemology (of the sort that implies [3]) or that Aristotle rejects metaphysical

Digitized by Arya Samaj Foundation Chennai and eGangotri

Even if Plato did hold a nonperspectival position, I do not see that it would commit him to the conclusions about value which Nussbaum derives from it. She thinks that the nonperspectival position commits Plato to denying that any relative value is true value at all (156); but she seems to rely on a confusion about relativity. What is healthy for a human being depends on a human being's nature, including human imperfections. But a more perfect being can recognize these facts, and can therefore see what is really healthy for a human being. Taking up a "nonanthropocentric" or "god's-eye" or "nonperspectival" point of view does not mean that we cannot recognize values that are valuable because of imperfections.

For these reasons, Nussbaum's claims about Plato on value contribute nothing to the understanding of the *Republic*.

According to Nussbaum, the *Phaedrus* rejects the view of the *Republic*. ⁴ She argues: (a) The "early" *Phaedrus* (Lysias' speech and Socrates's first speech) express the views of the *Republic*. (b) The "later" *Phaedrus* (Socrates's second speech) corrects the earlier part, and is incompatible with the *Republic* and the *Symposium*.

Against (a), I think Nussbaum misses some differences between the early *Phaedrus* and the *Republic*: (1) The early *Phaedrus* advises a young man to avoid sexual intercourse with any man who feels passionate love for him, and to engage in it only with a nonpassionate "nonlover" whose support and patronage will improve the young man's own prospects in life.⁵ The only alleged parallel that Nussbaum cites from the *Republic* is the probable endorsement of "some comfortable and non-passionate homosexual relations" (210). Once we distinguish (as Nussbaum does not) this recreational view of sex

Nussbaum suggests that this strictly instrumental and mercenary attitude to sexual relations is, in the historical and social circumstances envisaged by Plato, quite reasonable; she claims that feminists would also recommend it to a young woman in relation to some of her male colleagues in her profession (207f).

⁴ She claims that Plato's homosexual passion for Dion explains (at least partly) his change of mind (230), since the very writing of an epigram on Dion's death allegedly conflicts with Plato's previous prohibition on laments for the dead (*Republic* 387de). Nussbaum needs to prove: (1) Plato never had any sexual passion of equal intensity before his passion for Dion. (2) The epigram violates the prohibition in the *Republic*. (3) Plato regarded the provisions for the ideal state as moral principles for the conduct of his own life. (4) Plato never did anything that violated his moral principles. We lack any historical evidence of (1), (3), and (4), and Nussbaum does not argue for them. Her argument for (2) is obscure. It is easy to agree that the epigram expresses intense feeling; but it also strikes me as quite calm and restrained, not in itself evidence of the lamentation prohibited in the *Republic* (see the examples given in 388a–c, sharply different from Plato's epigram). On these grounds, I would be inclined to dismiss Nussbaum's excursion into psychobiography.

Digitized by Arya Samaj Foundation Chennai and eGangotri from the mercenary view assumed in the early *Phaedrus*, we see that there is no genuine parallel. (2) The early *Phaedrus* recognizes two parts of the soul (237d5–9), not the three parts recognized both in the *Republic* and in the later *Phaedrus*; one part—appetite—is innate, and the other—belief aiming at the best—is acquired later. In the *Republic*, by contrast, both rational and nonrational parts include their own specific types of desire (*epithumia*; 431b9–d2, 580c7–8). (3) The only function ascribed to practical reason, in the early *Phaedrus*, is to arrange the orderly satisfaction of nonrational desires. The *Republic* condemns this purely instrumental view as the outlook of the oligarchic man (553d). It is quite inept of Plato to introduce this contrast (not mentioned by Nussbaum) if he means to identify the views of the early *Phaedrus* with those of the *Republic*.

Against (b), I think Nussbaum is wrong to contrast the later Phaedrus with the Republic on some points. (1) She argues that the nonrational parts of the soul are treated more favorably in the later Phaedrus than in the Republic. She merely mentions and dismisses an apparent counterexample—the treatment of the thumos in the Republic (155, 456n5). She greatly exaggerates the brutish character of appetites, and she takes no account of Plato's demand for friendship and concord between the parts. (2) She claims that the role of erôs and madness in the Phaedrus conflicts with the Republic, because it is the domination of the rational part by the nonrational. But this is not Plato's view; the relevant type of madness is a desire of the rational part of the soul (see, e.g., 253e5-6), not a nonrational desire. (3) Equally, the Republic agrees with the Phaedrus in ascribing erôs (described in very similar sexual terms) to the rational part of the soul (see Republic 490ab). (4) The Phaedrus, according to Nussbaum, conflicts with the Symposium as well as the Republic, in so far as it values love of individual persons for themselves, not simply as instrumental means to awareness of the Forms. Nussbaum assumes that the Socratic and Platonic "One over Many" principle, claiming that beauty is a single property, implies the qualitative identity and interchangeability of all instances of beauty (179; Symposium 210a8-b4).6 Her assumption is surprising; for the "One over Many" principle is consistent with belief in different, noninterchangeable species of beauty (if 'animal' is univocal, it does not follow that there are no different species of animal).

Nussbaum's arguments, therefore, do not warrant any confidence in her views about the *Phaedrus*.

⁶ She appeals to the "ocean of beauty," 210d4 (180). Such a metaphor by itself hardly implies qualitative identity.

According to Nussbaum, Aristotle rejects Plato's view and accepts the tragedians' view on the vulnerability of the good life to external circumstances. I do not think Aristotle is as far from Plato as Nussbaum makes him seem.

Plato disagrees with Homer and the tragedians in so far as he claims that the good person is better off than anyone else, no matter what else happens to him; and Nussbaum gives no reason to suppose that Aristotle disagrees with Plato on this claim. He urges that the happy person can never become wretched (athlios), "because he will never engage in hateful and base actions" (Nicomachean Ethics 1100b34-5), and will always (aei, 1101b2) do what is best in the circumstances. In denying that the happy person could ever become wretched, Aristotle rejects a universal assumption of the tragedians (see, e.g., Sophocles, Oedipus Tyrannus 1196-1206). In Nussbaum's paraphrase, however, 'always' is abruptly and unjustifiably transformed into 'often'. She obscures the extent of Aristotle's disagreement with the tragic outlook and the extent of his agreement with Plato.7

Nussbaum's view makes it hard for her to see why Aristotle disapproves of tragedies about disaster befalling a good person as "revolting" (Poetics 1452b34-6), and why he requires the "tragic hero" to be less than completely good. She does not even mention the first claim, and she arbitrarily assumes that the second claim rules out only the person of superhuman, godlike virtue (387).

I do not have the space for a proper discussion of Nussbaum's concurrent argument defending a prominent role for literature in moral philosophy.8 Literature may evoke a particular sort of emotional response in the reader, and this "can sometimes be not just a means to practical knowledge, but a constituent part of the best sort of recognition or knowledge of one's practical situation" (15f). Nussbaum seems to believe that, since (i) moral philosophers should take literature seriously, it follows that (ii) they should regard literary

The literary works that Nussbaum mainly discusses are works of fiction, especially some Greek tragedies and some novels of Henry James. She seems to assume that only works of fiction—as opposed to, e.g., some works of history and biography—can perform the appropriate philosophical functions. I see no reason to accept the assumption.

⁷ One might well argue that Aristotle's claim about the invulnerability of virtue is inconsistent with some of his other claims about virtue and activity (though Nussbaum's arguments about the vulnerability of virtue (334f) are unconvincing). But, even if he ought to concede the vulnerability of virtue as a whole, he might still claim that the part of virtue which consists in a firm disposition to find the best and finest thing in the circumstances is invulnerable to external circumstances.

Digitized by Arya Samaj Foundation Chennai and eGangotri writers and texts as seeking and finding the truth, and (iii) they should write philosophical treatises that include the emotion-arousing features of literature.

Her belief that (i) implies (ii) underlies her objections to modern scholars who suggest that Aeschylus' assumptions about moral obligation are inconsistent; for she suggests that they have "assailed" Aeschylean tragedy as "morally primitive" (25), and therefore as not "helpful to us in exploring our modern beliefs about goodness of choice" (26). Her suggestion is dubious. Moral assumptions that are jointly inconsistent may nonetheless seem individually compelling. and a tragedy might make us feel their compelling character.

Although Nussbaum initially accepts (iii), and criticizes Iris Murdoch for rejecting it (16f), she recognizes much later that Aristotle's moral philosophy satisfies (i) but not (iii) (391-393). Her grounds for (iii) remain obscure (393f). She might argue that some claim to moral knowledge, of the sort that is appropriately sought by a moral philosopher, is justifiable only by appeal to the reader's emotional response to a particular fictional case presented by literary methods. As far as I can see, she wisely stops short of such a defense of (iii).

Nussbaum suggests that among the Greeks "nobody thought of their [sc. the epic and tragic poets'] work as less serious, less aimed at truth, than the speculative prose treatises of historians and philosophers" (12; cf. 124). This claim about "the Greeks" is falsified by Plato's view that tragedy is a form of rhetoric, which aims at pleasing the audience, not at the truth (Gorgias 502e). Plato's view was no doubt controversial; but Nussbaum gives no reason to suppose that the view she mentions was even approximately unanimous.9 It is striking (and not remarked by her) that no part of Aristotle's estimate of tragedy rests on any claim that Sophocles, say, was pursuing the truth, or that his views reliably state moral truths.

Nussbaum's three problems arise from Attic tragedy in so far as it describes the relevant sorts of conflicts. But I doubt whether, as she claims (8), they are all problems for the tragedians. For the tragedians seem not to think there is any question about whether the conflicts happen; they assume that all the conflicts happen, and display how they happen. Greek philosophers, however, think there

⁹ Nussbaum quotes Heracleitus criticizing Hesiod along with Pythagoras, Xenophanes, and Hecataeus, as illustrations of Heracleitus' point that "information about many things does not teach understanding"; and she claims that "all are criticized as searchers for understanding" (199). The control of the hor claims Heracleitus as searchers for understanding (199). criticized as searchers for understanding" (123). I see no basis for her claim; Heracleitus does not say a word about why Hesiod collected all the information he possessed, and definitely does not suggest that Hesiod was looking for understanding.

is a serious question about whether the conflicts actually arise once we straighten out our moral principles, and about whether they arise in just the ways assumed in tragedy. In claiming that philosophy is continuous with tragedy, Nussbaum overlooks this apparently important discontinuity. VI

Nussbaum's treatment of the texts she discusses displays a pattern of distortion and omission whose cumulative effect is to conceal evidence that appears to conflict with her general view. Readers will be well advised to check her translations and paraphrases of Greek texts and her descriptions of the views of modern philosophers and scholars. 10 The faults in the argument of this book justify skepticism about its main claims.

T. H. IRWIN

Cornell University

An example of Nussbaum's treatment of modern scholars is her account of "Kant-influenced commentators such as Sir David Ross and H. H. Joachim," according to whom "the gifts and reversals of fortune can never diminish eudaimonia" (329). I do not see where Joachim expresses this view [Nicomachean Ethics (New Vorl.)] (New York: Oxford, 1951), 58f]. The reference to Ross is to the wrong work. It is probably intended to refer to his Aristotle (London: Methuen, 1923), p. 192, where he does not express the view that Nussbaum attributes to him. Nussbaum's insinuation the tion that the influence of Kant suggested Ross's and Joachim's alleged interpretation to them is unwarranted.

¹⁰ I have commented above on Republic 387d-388a, 504; Gorgias 495bc; Nicomachean Ethics 1101b2. See also 261 ('makes conjectures' for peirastikê in Metaphysics 1004b25-6), 305 ('ultimately simple' for prôtôn and 'external justification' for logos, in Nicomachean Ethics 1143a36-b1). In quoting the last passage, Nussbaum does not suggest that her rendering is meant as anything other than a translation; in a footnote she says she follows Wiggins's translation-cum-paraphrase to some extent (492), but she does not say she is only offering a paraphrase. It matters which she is offering; for if readers suppose that it is a translation and that something corresponding to 'external' appears in the Greek, they may wonder how Nussbaum's interpretation of logos could ever be controversial, when in fact it is highly controversial. Again, she claims that "choice (prohairesis) is described as an ability that is on the borderline between the intellectual and the passional, partaking of both natures" (307); but she cites passages that say prohairesis involves both reason and desire, orexis, not passion, pathos. Aristotle makes it clear that the relevant sort of desire is rational desire, boulesis, which is not a pathos.

Digitized by Arya Samaj Foundation Chennai and eGangotri *Psychosemantics*. JERRY FODOR. Cambridge: MIT Press, 1987. 173 p. \$19.95.

In Word and Object, W. V. O. Quine1 acknowledged the "practical indispensability" in daily life of the intentional idioms of belief and desire, but disparaged such talk as an "essentially dramatic idiom" rather than something from which real science could be made in any straightforward way (ibid., p. 219). Many who agree on little else have agreed with Quine about this, and have gone on to suggest one or another indirect way for science to accommodate folk psychology: Wilfrid Sellars, Donald Davidson, Hilary Putnam, Richard Rorty. Stephen Stich, Patricia Smith Churchland, Paul Churchland, Stephen Schiffer, and myself, to name a few. This fainthearted consensus is all wrong, according to Fodor, whose new book is a vigorous—even frantic—defense of what he calls Intentional Realism: beliefs and desires are real, causally involved, determinately contentful states. "We have no reason to doubt," Fodor says, "that it is possible to have a scientific psychology that vindicates commonsense belief/desire explanation" (16).

The first of the four chapters motivates the intricate set of arguments to follow with a spirited paean to the powers of commonsense psychology, but it adds up to less than its rhetoric leads one to expect. The opposition, after all, has not been unimpressed with what Quine called the practical indispensability of the intentional idiom; I, for one, have tirelessly sung the praises of what can be done from the intentional stance, and even Paul Churchland,2 the most radical of the critics, has acknowledged, and even offered a metaphysical analysis of, the systematic predictive power of propositional attitude attributions. What we doubters have not yet seen is the argument from the practical indispensability of propositional attitude talk to Fodor's hard-core realism, which this chapter should presumably provide. It stops short, however, with a tactical proposal: "we can't give them up because we don't know how to. So maybe we had better try to hold on to them" (10; see also the concession in footnote 2, p. 155). What then follows is a reasoned catalogue of what the hard-core realist is committed to: in the main, the Representational Theory of Mind, which claims there are "entities which —like the attitudes—are both semantically evaluable and etiologically involved" (26). Something has slipped, however. In which camp

¹ Cambridge: MIT, 1960.

0022-362X/88/8507/0384\$00.50

© 1988 The Journal of Philosophy, Inc.

² Scientific Realism and the Plasticity of Mind (New York: Cambridge, 1979), pp. 101-107.

would Fodor place someone who was a realist about entities which met these two conditions but were not *enough* like the propositional attitudes to constitute a (realist) vindication of commonsense psychology?

Never mind. Fodor's tactical proposal is well worth adopting in the spirit of scientific opportunism: try for the boldest, simplest, most satisfying theory first, and see where it leads you. Recent history offers a nice example on Fodor's side: the proposal that we should be straightforward realists about Mendelian genes and treat them as tokens of expressions in the language of inheritance, straightforwardly realized in hunks of DNA, has been a vastly informative oversimplification. If, now that we understand the underlying mechanisms better, we see that there are theoretically important mismatches between the language of "bean-bag genetics" and the molecular and developmental details—mismatches serious enough to suggest that, all things considered, there do not turn out to be genes (classically understood) after all—this is an understanding which would have been well-nigh inaccessible without the ladder of realist oversimplification. So we can proceed with the exploration of the implications of intentional realism, secure in the recognition that we can always back off to a position of instrumentalism (like the current instrumentalism about classical genetics) when we have learned more.

The second chapter deals with the oft-explored problems posed by the Twin Earth thought experiments of Tyler Burge, Putnam, and others. If psychology is going to be science, it had better not posit mysterious action-at-a-distance, so a principle of the "supervenience" of the psychological on the physiological must be honored: the brains of organisms differ whenever their minds differ (30). At the same time, if psychology is going to deal with real, determinate content, it must honor the principle that "extension constrains content." The only way to meet both conditions at once is to distinguish notions of narrow and broad content. Narrow content is supervenient on the organism's internal, causal states, and broad content is then held to be a function of narrow content relativized to environmental context. Plunking an organism into different contexts while holding its brain constant cannot change its narrow content, but can in principle alter the broad content of its psychological states. This is a proposal that has often been made before, for the same reasons (see, for instance, my defense of "notional attitude psychology"),3

³ "Beyond Belief," in Andrew Woodfield, ed., Thought and Object: Essays on Intentionality (New York: Oxford, 1982), pp. 1-95.

Digitized by Arya Samaj Foundation Chennal and eGangotri

but Fodor adds novel intricacies in its defense to counter the increasingly intricate objections that have been spawned in subsequent years. (If Fodor wants to fight this fight, I will happily hold his coat. suppressing, on this occasion, the doubts I have about some of his background assumptions.)

It is in chapter three that Fodor begins to distinguish himself sharply from those he takes to be the enemies of scientific psychology. When everything is held to merge with everything else, when there are no clean joints at which to carve nature, science tends to wind down to a lazy halt: holism as the heat death of science. (The theory of "blending inheritance" which stymied Darwin might be cited as a telling example.) So Fodor sets himself up as the guardian of good crisp boundaries, and hence the enemy of all relativism. pragmatism, and various other unspeakable but fashionable doctrines, banded together under the black flag of Meaning Holism. Science is nothing if not "naturalistic," however, and, when one tries to naturalize psychology, it begins to look more and more biological. and hence teleological, and hence functionalistic; its crisp traditional categories—in particular, propositional attitudes—begin to slide toward the murky relativity of functional interpretation, and beyond: to the indeterminacy of radical interpretation.

Some version or other of meaning holism is the natural resting place of this slide, but it "really is a crazy doctrine," Fodor says (60). (This sort of statement should ring alarm bells in any author. What are the odds that all those worthy opponents out there are crazy enough to hold the crazy doctrine one is arguing against?) According to meaning holism, the content of any particular belief depends on what else you believe (now is that not the silliest thing you ever heard?). Fodor rebuts various arguments for meaning holism (unpersuasively, in my opinion, for reasons that would not fit in the space allotted me), but in the process he gives some hostages to the opposition.

First, he distinguishes sharply between two brands of functionalism, and endorses only the weaker version:

. . . all you need is the claim that being a belief is a matter of having the right connections to inputs, outputs and other mental states. What you don't need-and what the philosophical motivations for Psychofunctionalism therefore do not underwrite—is the much stronger claim that being the belief that p, being a belief that has a certain content is a matter of having the right connections to inputs, outputs, and other mental states (69).

So Fodor ends up defending an eviscerated functionalism, and disagreeing with those of us, from Sellars to the present, who in one

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

way or another take "functional role semantics" to be the chief beauty of functionalism. For Fodor, content is not a function of function after all. Or, at any rate, it is not much of a function of function, for Fodor must yield another hostage: functional role "is a marginal—a not very important—determinant of meaning" (72). If content is not (primarily) a matter of functional role, what is it a matter of? Denotation. According to his alternative denotational semantics, "concepts are individuated by reference to the properties they express, thoughts by the states of affairs they correspond to, and so forth" (91). Independently of any functional role that a concept or mental representation might play within an interanimated set of its fellows, it can, as a matter of brute metaphysical fact, denote (or express) a property, thanks to its causal (but not functional) link to that property.

The task then falls to him, in the last chapter, to make good on this by defining in nonsemantic, nonfunctional, nonteleological terms the conditions under which a mental representation expresses a property. In the end, after successive elaborations and improvements, we are offered the Slightly Less Crude Causal Theory of Content: a "sufficient condition for 'A's to express A is that it's nomologically necessary" that (1) "All instances of A's cause 'A's when (i) the A's are causally responsible for psychophysical traces to which (ii) the organism stands in a psychophysically optimal relation"; and (2) "If non-A's cause 'A's, then their doing so is asymmetrically dependent upon A's causing 'A's" (126).

What with its mention of intact organisms and psychophysically optimal conditions, one can certainly doubt that this account succeeds in avoiding appeal to teleological assumptions. Even if, by some marvel of circumlocution, it can be held to skirt those shoals, such exercises, like the desperate definitional evasions of the behaviorists, do not so much defeat as pay homage to the contrary insight. Moreover, Fodor's account forces him to recant his analysis of the Twin Earth cases (if I have correctly traced out the requisite counterfactuals—Fodor gives the reader scant advice on this). Suppose some stray XYZ causes Fodor's Mentalese "water" symbol to light up; surely its doing this is not asymmetrically dependent on H₂O doing this; so 'water' for Fodor means H₂O-or-XYZ; Fodor and his Doppelgänger are broad-content twins after all. So far as I can see, all the escape routes from this conclusion are blocked by conclusions even less palatable to Fodor.

In any event, other pesky counterexamples suggest themselves, and Fodor himself acknowledges that his account still has some problems he does not yet know how to solve. In the meantime, he usefully draws attention to some of the implications he will invite us

to swallow once he has sorted out those problems. Consider the young child who uses the word 'father', recognizes his father, etc., but has so far given no sign of having the concepts MALE or PAR-ENT. Does he have the concept FATHER? The meaning holist is free to say that the child only gradually comes to approximate having the adult concept of FATHER, as he acquires more relevant beliefs.4 but, for Fodor, the meaning atomist, there is a deep (and quite possibly unknowable) fact about what concept the child actually has. It may, unwittingly, already have the concept of FATHER, in spite of its disappointing responses to a variety of presumably telling tests: "you can have the concept MALE PARENT without having access to its internal structure" (161). This independence of meaning from function on Fodor's view is even more striking in an objection he raises against himself: "'On your view, entertaining (as it might be) the thought that three is a prime number could constitute an entire mental life?' This too is satisfactory as a reductio ad absurdum only on the assumption that its conclusion is false" (89). I do not think these implications show that meaning atomism is crazy, but as Fodor says, in a similar connection, "chaque à son rocker" (sic, p. 124).

In an appendix, "Why There Still Has to Be a Language of Thought," Fodor offers three arguments for why the internal, semantically evaluable, etiologically involved states of intentional realism—the beliefs and desires—have to be organized in a language-like system. He is right to have isolated these arguments in an appendix, for they are independent of his main argument; if you have bought the rest of his case, they do indeed sharpen his account, but, if you have not, there are ways of recasting them all to conform to your unrepentant meaning holism or even instrumentalism.

The convolutions of argument found in *Psychosemantics* make it a difficult book to review fairly in short compass. The argument strategy adopted by Fodor throughout the book is to emphasize small disagreements and generally force all fence sitters to jump one way or the other. Then, by a series of nested arguments designed to corral the obstreperous dogies, he can end up having fenced off the only defensible territory. There are several weaknesses in such a strategy. First, a single gap in the fence is all it takes to escape Fodor's roundup, and so the argument tends to degenerate into fence-busting, fence-mending squabbles about minutiae—very important minutiae, Fodor would insist, but I find it hard to believe that this is the way to make progress. Moreover, this emphasis on amplifying disagreements can backfire in the end, when it is time for

⁴ Cf. my Content and Consciousness (Boston: Routledge & Kegan Paul, 1969), p. 183.

Fodor to make his ecumenical concessions, as he does, to verificationism or procedural semantics, to behaviorism or associationism, and the other heresies of previous chapters. It turns out that the crazies were not so crazy after all. "Whenever I tell this story to Granny, she grins and rocks and says 'I told you so'" (118). "All right, Granny, have it your way . . . Only do stop rocking; you make me nervous" (122).

Finally, Fodor is too wise to think his series of arguments can flat disprove the claims of the opposition, so time and again he resorts to claims about shifting the burden of proof, begging the question, outsmarting by embracing the conclusions of reductions, and other exploitations of the rules of the game. The book is a tireless exercise of that philosopher's pastime, burden-tennis. Burden, burden, who has the burden of proof now? Fodor mostly plays solitaire burdentennis, against an imaginary opponent often personified as Granny or Aunty, which permits him to express the opposition view in terms that suit his rebuttal, without having to address the issue of whether this is a sympathetic rendering of any real opponent's claims.

All this fancy footwork and shadowboxing leads Fodor into an increasingly tight place, which reminds Aunty, bless her heart, of the time Winnie-the-Pooh got himself similarly stuck when trying to leave Rabbit's house. Pooh's idea was to fast for a week, while Christopher Robin read him a book. Fodor, in his increasingly desperate effort to free himself from the house of Gavagai, has written himself a book, and while it may not be to everyone's taste, Aunty thinks *Psychosemantics* is just the sort of book Pooh had in mind: "a Sustaining Book, such as would help and comfort a Wedged Bear in Great Tightness."

DANIEL C. DENNETT

Tufts University

⁵ A. A. Milne, Winnie-the-Pooh (New York: Dutton, 1926), p. 30.

NEW BOOKS

ACKERMANN, ROBERT JOHN: Wittgenstein's City. Amherst: Massachusetts UP, 1988. xiii, 267 p. Cloth \$25.00, paper \$12.95.

ANNAS, JULIANA: Oxford Studies in Ancient Philosophy: vol. V. New York: Oxford, 1988. 257 p. Cloth \$54.00, paper \$26.00.

ASHBAUGH, ANNE FREIRE: Plato's Theory of Explanation: A Study of the Cosmological Account in the Timaeus. Albany: SUNY, 1988. vii, 195 p. Cloth \$34.50, paper \$12.95.

BAKER, LYNNE RUDDER: Saving Belief: A Critique of Physicalism. Princeton: University Press, 1988. xii, 177 p. \$19.95.

BARASH, JEFFREY ANDREW: Martin Heidegger and the Problem of Historical Meaning. Boston: Nijhoff, 1988. xviii, 322 p. \$68.50.

BENCIVENGA, ERMANNO: Kant's Copernican Revolution. New York: Oxford, 1987. viii, 262 p. \$32.50.

BEROFSKY, BERNARD: Freedom From Necessity: The Metaphysical Basis of Responsibility. New York: Routledge & Kegan Paul, 1988. viii, 231 p. \$27.50.

BIRD, GRAHAM: William James. New York: Methuen, 1987. ix, 221 p. \$39.95. BLUMENBERG, HANS: Die Sorge geht über den Fluss. Frankfurt/Main: Suhrkamp, 1987. 222 p.

BORZESZKOWSKI, H. H. VON, and H. J. TREDER: The Meaning of Quantum Gravity. Dordrecht: Kluwer, 1987. vii, 132 p. \$48.00.

BRODY, BARUCH: Life and Death Decision Making. New York: Oxford, 1988. xii, 250 p. \$27.95.

BRODY, HOWARD: Stories of Sickness. New Haven: Yale, 1987. xiii, 210 p. \$27.50.

BROWN, DAVID: Continental Philosophy and Modern Theology. New York: Blackwell, 1987. vii, 250 p. \$49.95.

BROWN, LES: Conservation and Practical Morality: Challenges to Education and Reform. New York: St Martin's, 1988. ix, 222 p. \$32.50.

BROWN, ROBERT: Analyzing Love. New York: Cambridge, 1987. vii, 127 p. \$24.95.

BRUNN, EMILY ZUM: St. Augustine: Being and Nothingness. New York: Paragon, 1987. xiii, 129 p. \$21.95.

CALLINICOS, ALEX: Making History. Ithaca: Cornell, 1988. viii, 275 p. \$39.95.

CARLSNAES, WALTER: Ideology and Foreign Policy: Problems of Comparative Conceptualization. New York: Blackwell, 1987. xi, 234 p. \$39.95.

CARTWRIGHT, RICHARD: Philosophical Essays. Cambridge: MIT, 1987. xxiii, 266 p. \$25.00.

CAVELL, STANLEY: Disowning Knowledge in Six Plays of Shakespeare. New York: Cambridge, 1987. x, 226 p. Cloth \$34.50, paper \$10.95.

CELA-CONDE, CAMILO: On Genes, Gods and Tyrants. Boston: Reidel, 1987. xii, 201 p. \$59.00.

CHAITIN, G. J.: Algorithmic Information Theory. New York: Cambridge, 1987. iv, 175 p. \$29.95.

COCCHIARELLA, NINO B.: Logical Investigations of Predication Theory and the

- Problem of Universals. Atlantic Highlands, NJ: Humanities, 1986. 272 p. \$50.00. Logical Studies in Early Analytic Philosophy. Columbus: Ohio UP, 1987. 296 p. \$30.00.
- corrington, robert s.: The Community of Interpreters: On the Hermeneutics of Nature and the Bible in the American Philosophical Tradition. Macon, GA: Mercer, 1988. xiii, 109 p. \$24.95.
- CRUZ, FEODOR F.: John Dewey's Theory of Community. New York: Lang, 1988. 213 p. \$33.50.
- DANIELS, NORMAN: Am I My Parents' Keeper? An Essay on Justice between the Young and Old. New York: Oxford, 1988. x, 194 p. \$19.95.
- DAVIDSON, NICHOLAS: The Failure of Feminism. Buffalo: Prometheus, 1988. 392 p. \$24.95.
- DONAGAN, ALAN: Choice: The Essential Element in Human Action. New York: Routledge & Kegan Paul, 1987. x, 197 p. \$29.95.
- DREAR, PETER: Mersenne and the Learning of the Schools. Ithaca: Cornell, 1988. xiii, 264 p. \$28.50.
- ENGLEBRETSEN, GEORGE: The New Syllogistic. New York: Lang, 1987. xii, 322 p. \$41.50.
- FEIBLEMAN, JAMES KERN: The Destroyers: The Underside of Human Nature. New York: Lang, 1987. 187 p. \$26.00.
- FERRARI, G. R. F.: Listening to the Cicadas: A Study of Plato's Phaedrus. New York: Cambridge, 1987. xiii, 293 p. \$39.50.
- FIELD, J. v.: Kepler's Geometrical Cosmology. Chicago: University Press, 1988. xx, 243 p. \$37.50.
- FLEW, ANTONY, and GODFREY VESEY: Agency and Necessity. New York: Blackwell, 1987. 184 p. Cloth \$34.95, paper \$14.95.
- FLEW, ANTONY: The Logic of Mortality. New York: Blackwell, 1987. ix, 200 p. \$45.00.
- FRANZEN, TORKEL: Provability and Truth. Stockholm: Almqvist & Wiksell, 1987. 81 p.
- GARFIELD, JAY L.: Belief in Psychology: A Study in the Ontology of Mind. Cambridge: MIT, 1988. x, 168 p. \$19.95.
- GOLDSTEIN, ROBERT D.: Mother-love and Abortion: A Legal Interpretation. Berkeley: California UP, 1988. xii, 240 p. \$19.95.
- GRAHAM, DANIEL W.: Aristole's Two Systems. New York: Oxford, 1988. xiii, 359 p. \$66.00.
- GUYER, PAUL: Kant and the Claims of Knowledge. New York: Cambridge UP, 1987. xiii, 482 p. Cloth \$59.50, paper \$19.95.
- GYEKYE, KWAME: An Essay on African Philosophical Thought: The Akan Conceptual Scheme. New York: Cambridge, 1987. xv, 246 p. \$34.50.
- HACKER, P. M. S.: Appearance and Reality. New York: Blackwell, 1987. x, 243 p. \$45.00.
- HALL, ROBERT T.: Emile Durkheim: Ethics and the Sociology of Morals. New York: Greenwood, 1987. x, 234 p. \$37.95.
- HALLETT, GARTH L.: Language and Truth. New Haven: Yale, 1988. viii, 234 p. \$21.50.

- HARPHAM, GEOFFREY GALT: The Ascetic Imperative in Culture and Criticism. Chicago: University Press, 1987. xvii, 325 p. \$29.95.
- HARVEY, STEVEN: Falaquera's Epistle of the Debate: An Introduction to Jewish Philosophy. Cambridge: Harvard, 1987. xvi, 155 p. Cloth \$22.50, paper \$14.00.
- HILEY, DAVID R.: Philosophy in Question: Essays on a Pyrrhonian Theme. Chicago: University Press, 1988. ix, 207 p. \$24.95.
- HOULGATE, LAURENCE D.: Family and State: The Philosophy of Family Law. Totowa, NJ: Rowman & Littlefield, 1988. xvi, 207 p. \$33.50.
- HURFORD, JAMES R.: Language and Number: The Emergence of a Cognitive System. New York: Blackwell, 1987. xi, 322 p. \$49.95.
- JONES, RICHARD H.: Science and Mysticism. Cranbury, NJ: Associated, 1986. 266 p. \$20.00.
- KAINZ, HOWARD P.: Paradox, Dialectic, and System: A Contemporary Reconstruction of the Hegelian Problematic. University Park: Penn State, 1988. ix, 137 p. \$20.00.
- KATZ, LEO: Bad Acts and Guilty Minds: Conundrums of the Criminal Law. Chicago: University Press, 1987. xii, 343 p. Cloth \$45.00, paper \$14.95.
- KAVKA, GREGORY S.: Moral Paradoxes of Nuclear Deterrence. New York: Cambridge, 1987. 243 p. \$34.50.
- KENNY, ANTHONY: The Heritage of Wisdom: Essays in the History of Philosophy. New York: Blackwell, 1987. ix, 179 p. \$24.95. Reason and Religion: Essays in Philosophical Theology. New York: Blackwell, 1987. x, 182 p. \$24.95.
- KURTZ, PAUL: Forbidden Fruit: The Ethics of Humanism. Buffalo: Prometheus, 1988. 266 p. \$19.95.
- LEMOS, RAMON M.: Metaphysical Investigations. Cranbury, NJ: Fairleigh Dickinson, 1988. 290 p. \$35.00.
- MACINTYRE, ALASDAIR: Whose Justice? Which Rationality? Notre Dame: University Press, 1988. xi, 410 p. \$22.95.
- MACKEY, JAMES P.: Modern Theology: A Sense of Direction. New York: Oxford, 1987. 200 p. Cloth \$26.00, paper \$8.95.
- MARENBON, JOHN.: Later Medieval Philosophy (1150-1350): An Introduction. New York: Routledge & Kegan Paul, 1987. xi, 230 p. \$35.00.
- MARGOLIS, JOSEPH: Science Without Unity: Reconciling the Human and Natural Sciences. New York: Blackwell, 1987. xxii, 470 p. \$39.95.
- MAY, KEITH M.: Nietzsche and Modern Literature: Themes in Yeats, Rilke, Mann, and Lawrence. New York: St. Martin's, 1988. ix, 175 p. \$29.95.
- MILLER, PETER: Domination and Power. New York: Routledge & Kegan Paul, 1988. viii, 269 p. \$37.50.
- MORRIS, PETER: Power: A Philosophical Analysis. New York: St. Martin's, 1988. 266 p. \$35.00.
- MURDOCH, DUGALD: Niels Bohr's Philosophy of Physics. New York: Cambridge, 1987. x, 294 p. \$59.50.
- NORRIS, CHRISTOPHER: Derrida. Cambridge: Harvard, 1988. 271 p. Cloth \$25.00, paper \$8.95.
- NORTON, BRYAN G.: Why Preserve Natural Variety? Princeton: University Press, 1987. xiii, 281 p. \$30.00.

Digitized by Arya Samaj Foundation Chennai and eGangotri

O'DALY, GERARD: Augustine's Philosophy of Mind. Berkeley: California UP, 1987. xii, 241 p. \$35.00.

ORRU, MARCO: Anomie: History and Meanings. Boston: Allen & Unwin, 1987. xii, 209 p. \$34.95.

OUTHWAITE, WILLIAM: New Philosophies of Social Science: Realism, Hermeneutics, and Critical Theory. New York: St. Martin's, 1988. viii, 137 p. \$35.00.

PAGET, MARIANNE A.: The Unity of Mistakes: A Phenomenological Interpretation of Medical Work. Philadelphia: Temple, 1988. xiii, 183 p. \$24.95.

PAPINEAU, DAVID: Reality and Representation. New York: Blackwell, 1987. xix. 243 p. \$39.95.

PEARS, DAVID: The False Prison: A Study of Wittgenstein's Philosophy. New York: Oxford, 1987. viii, 202 p. Cloth \$34.50, paper \$12.95.

PIVCEVIC, EDO: The Concept of Reality. Wolfeboro, NH: Longwood, 1987. ix, 296 p. \$30.00.

PORPORA, DOUGLAS V.: The Concept of Social Structure. New York: Greenwood, 1987. 159 p. \$37.95.

PRADO, C. G.: The Limits of Pragmatism. Atlantic Highlands, NJ: Humanities, 1987. ix, 191 p. \$39.95.

PRIEST, GRAHAM: In Contradiction: A Study of the Transconsistent. Boston: Nijhoff, 1987. xv, 279 p. \$79.00.

PRONKO, N. H.: From AI to Zeitgeist: A Philosophical Guide for the Skeptical Psychologist. New York: Greenwood, 1988. xvi, 265 p. \$45.00.

RAPACZYNSKI, ANDRZEJ: Nature and Politics: Liberalism in the Philosophies of Hobbes, Locke and Rousseau. Ithaca: Cornell UP, 1987. xi, 302 p. \$29.95.

RECANATI, FRANCOIS: Meaning and Force: The Pragmatics of Performative Utterances. New York: Cambridge, 1988. x, 278 p. \$39.50.

REDNER, HARRY: The Ends of Science: An Essay in Scientific Authority. Boulder: Westview, 1988. xiv, 344 p. \$39.50.

ROBERTS, JULIAN: German Philosophy: An Introduction. Atlantic Highlands, NJ: Humanities, 1988. viii, 296 p. Cloth \$45.00, paper \$15.00.

ROLSTON, HOLMES III: Environmental Ethics: Duties to and Values in the Natural World. Philadelphia: Temple, 1988. xiii, 391 p. \$34.95.

ROSEN, STANLEY: Hermeneutics as Politics. New York: Oxford, 1987. 213 p.

RUSSELL, JAMES: Philosophical Perspectives on Developmental Psychology. New York: Blackwell, 1987. ix, 262 p. \$45.00.

RYCHLAK, JOSEPH F.: The Psychology of Rigorous Humanism. New York: University Press, 1988. 571 p. \$40.00.

SAJAMA, SEPPO, and MATTI KAMPPINEN: A Historical Introduction to Phenomenology. New York: Croom Helm, 1987. vii, 131 p. Cloth \$47.50, paper \$9.95.

SAMUELSON, WILLIAM G., and FRED A. MARKOWITZ: An Introduction to Philosophy in Education. New York: Philosophical Library, 1988. 190 p. \$12.95.

SAVILE, ANTHONY: Aesthetic Reconstructions. New York: Blackwell, 1988. ix, 254 p. \$45.00.

SEDDON, KEITH: Time: A Philosophical Treatment. New York: Croom Helm, 1987. 166 p. \$47.50.

SEN, AMARTYA: On Ethics and Economics. New York: Blackwell, 1987. 131 p. \$24.95.

SHKLAR, JUDITH N.: Montesquieu. New York: Oxford, 1987. viii, 136 p. Cloth 18.95, paper \$4.95. SHUSTERMAN, RICHARD: T. S. Eliot and the Philosophy of Criticism. New

York: Columbia, 1988. xii, 236 p. \$35.00.

SILVERMAN, HUGH J.: Inscriptions: Between Phenomenology and Structuralism. New York: Methuen, 1988. xvi, 390 p. \$29.95.

SIMONS, PETER: Parts: A Study in Ontology. New York: Oxford, 1987. xiii, 390 p. \$72.00.

SMITH, JOSEPH WAYNE: The Progress and Rationality of Philosophy as a Cognitive Enterprise. Brookfield, VT: Gower, 1988. vi, 301 p. \$68.50.

SORELL, TOM: Moral Theory and Capital Punishment. New York: Blackwell, 1988. viii, 172 p. Cloth \$29.95, paper \$15.95.

SPRIGGE, T. L. S.: The Rational Foundations of Ethics. New York: Routledge & Kegan Paul, 1988. x, 283 p. \$47.50.

SRZEDNICKI, JAN: The Democratic Perspective: Political and Social Philosophy. Boston: Nijhoff, 1988. vii, 376 p. \$79.50.

STERNFELD, ROBERT, and HAROLD ZYSKIND: Meaning, Relation, and Existence in Plato's Parmenides. New York: Lang, 1987. xiv, 187 p. \$29.50.

STROLL, AVRUM: Surfaces. Minneapolis: Minnesota UP, 1988. xi, 227 p. Cloth \$39.50, paper \$15.95.

SUGARMAN, SUSAN: Piaget's Construction of the Child's Reality. New York: Cambridge, 1987. vi, 258 p. \$34.50.

SUMNER, L. W.: The Moral Foundation of Rights. New York: Oxford, 1987. x, 224 p. \$45.00.

TEJERA, V.: Nietzsche and Greek Thought. Boston: Nijhoff, 1987. 157 p. \$34.50.

THOMSON, GARRETT: *Needs*. New York: Routledge & Kegan Paul, 1987. 143 p. \$37.50.

TRUSTED, JENNIFER: Inquiry and Understanding: An Introduction to Explanation in the Physical and Human Sciences. Atlantic Highlands, NJ: Humanities, 1987. ix, 171 p. Cloth \$29.95, paper \$9.95.

VOEGELIN, ERIC: In Search of Order. Baton Rouge: Louisiana State UP, 1987. 120 p. \$14.95.

WEISS, PAUL: Philosophy in Process, vol. X: April 15, 1984-January 18, 1986. Albany: SUNY, 1987. 393 p. Cloth \$49.50, paper \$19.50.

WINSTON, GORDON, and RICHARD TEICHGRAEBER III: The Boundaries of Economics. New York: Cambridge, 1988. xi, 122 p. \$27.50.

WOLLHEIM, RICHARD: Painting as an Art: The A. W. Mellon Lectures in the Fine Arts, 1984. Princeton: University Press, 1987. 384 p. \$45.00.

ZWEIBACH, BURTON: The Common Life: Ambiguity, Agreement, and the Structure of Morals. Philadelphia: Temple, 1988. xii, 216 p. \$24.95.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 8, AUGUST 1988

SANCTIFICATION, HARDENING OF THE HEART, AND FRANKFURT'S CONCEPT OF FREE WILL*

N a much-discussed paper,¹ "Freedom of the Will and the Concept of a Person," Harry Frankfurt presents an analysis of the self in terms of hierarchically ordered desires, and he uses his analysis to argue for a certain notion of freedom. His paper has generated considerable debate among philosophers interested in the concept of freedom and the related concept of autonomy. My own interest in Frankfurt's paper is primarily in applying his analysis of the self to problems in the philosophy of religion, especially to puzzles raised by the doctrine of sanctification and the notion of God's

* I am indebted to the following people for useful suggestions: Don Adams, John Christman, Philip Quinn, Bruce Russell, John Tyson, Peter van Inwagen, and Allen Wood. I owe a special debt to Harry Frankfurt, who corrected some important misunderstandings of his account, and to William Alston, whose objections helped me to think out some essential distinctions. Finally, I am particularly grateful to Norman Kretzmann, who gave me many helpful comments on earlier drafts.

For good summaries of the literature, see David Shatz, "Free Will and the Structure of Motivation," Midwest Studies in Philosophy, vol. x, Peter French, Theodore Uehling, Jr., and Howard Wettstein, eds. (Minneapolis: Minnesota UP, 1986), pp. 451-482; and Gary Watson, "Free Action and Free Will," Mind, XLVI (1987): 145-172. Besides the literature cited in those articles, I also found helpful the following articles (given in no particular order): Lawrence Haworth, Autonomy: An Essay in Philosophical Psychology and Ethics (New Haven: Yale, 1986), and "Autonomy and Utility," Ethics, xcv (1984): 5-19; William Rowe, "Two Criticisms of the A of the Agency Theory," Philosophical Studies, XLII (1982): 363-378; Richard Arneson, "Freedom and Desire," Canadian Journal of Philosophy, XV (1985): 425-448. 425-448; S. I. Benn, "Freedom, Autonomy and the Concept of a Person," Proceedings of the Land Dworkin "The ings of the Aristotelian Society, LXXVI (1975/6): 109-131; Gerald Dworkin, "The Concept of Autonomy," in Science and Ethics, R. Haller, ed. (Amsterdam: Rodopi, 1981) 1981), pp. 203-213; Irving Thalberg, "Socialization and Autonomous Behavior," in Studies 1979) in Studies in Action Theory, Robert Whittemore, ed. (New Orleans: Tulane, 1979), Pp. 21-37; Marilyn Friedman, "Autonomy and the Split-level Self," The Southern Journal of Philadelle (1997) (1997) The Importance of Free Journal of Philosophy, XXIV (1986): 19-35; Susan Wolf, "The Importance of Free Will", Mr. Will," Mind, xc (1981): 386–405.

This JOURNAL, LXVIII, 1 (January 14, 1971): 5–20. Other papers by Frankfurt which have some bearing on the issues discussed here are the following (given in no

Digitized by Arya Samaj Foundation Chennai and eGangotri. hardening hearts, although I think Frankfurt's analysis is remarkably fruitful for understanding a variety of religious claims and practices. In order to apply Frankfurt's views to issues in the philosophy of religion, however, it is important to reconsider his understanding of freedom and to refine his hierarchical analysis of the self. The resulting revised version of Frankfurt's account of freedom and the self is not vulnerable to the sorts of criticisms which have been leveled, quite correctly I think, against Frankfurt's original views. So in what follows I shall first consider Frankfurt's views of freedom and the self and suggest some revisions of them. Then I shall discuss the most important criticisms raised against Frankfurt's original position and argue that the revised Frankfurt account can be successfully defended against them. Finally, I shall show how that account can resolve some long-standing difficulties about sanctification and hardening of the heart.

I. FRANKFURT'S ACCOUNT

It is Frankfurt's view that the essence of a person is to be found in the structure of the will. He takes wants or desires to be the genus of acts of willing, or volitions, and he holds a volition to be an effective desire, which moves an agent all the way to action. According to Frankfurt, agents can have first-order desires and volitions—to do something—and also second-order desires and volitions—to have certain first-order desires. A person, on Frankfurt's view, is someone who has second-order desires and volitions. An agent who has no second-order volitions is "a wanton"; such an individual may be human but is not a person.

This analysis of the notion of a person is the basis for Frankfurt's account of freedom of will. The common conception of freedom as the ability to do what one wants to do, Frankfurt says, is best thought of as applying to freedom of action. Freedom of will can then be construed analogously as the ability to will what one wants to will, or the ability to have the sort of will one wants. On Frankfurt's view, in order to have freedom of will, an individual must meet the following

particular order): "Necessity and Desire," Philosophy and Phenomenological Research, XLV (1984): 1–13; "The Importance of What We Care About," Synthese, LIII (1982): 257–272; "The Problem of Action," American Philosophical Quarterly, XV (1978): 157–162; "Coercion and Moral Responsibility," in Essays on Freedom of Action, Ted Honderich, ed. (London: Routledge & Kegan Paul, 1973), pp. 65–86; "Three Concepts of Free Action II," in Proceedings of the Aristotelian Society, supp. vol. XLIX (1975): 113–125; "Alternate Possibilities and Moral Responsibility," this JOURNAL, LXVI, 23 (December 4, 1969): 828–839; "Identification and Externality," in The Identities of Persons, Amelie Rorty, ed. (Berkeley: California UP, 1976); "The Problem of Action," American Philosophical Quarterly, XV grateful to Frankfurt for letting me see this paper in typescript).

Digitized by Arya Samaj Foundation Chennai and eGangotri conditions: (1) he has second-order volitions, (2) he does not have first-order volitions that are discordant with those second-order volitions, and (3) he has the first-order volitions he has because of his second-order volitions (that is, his second-order volitions have, directly or indirectly, produced his first-order volitions; and if his second-order volitions had been different, he would have had different first-order volitions).³

This account might appear to imply that only a person who meets these strong criteria for freedom of the will is morally responsible for his actions, but Frankfurt is not committed to such a counter-intuitive view. He distinguishes between acting freely and having freedom of will when one acts. If a person has done what he wanted to do because he wanted to do it and the will by which he was moved when he did it was his own will, then he acted freely, even if he did not act with freedom of will. Assessments of moral responsibility, according to Frankfurt, should depend primarily on whether or not an agent acted freely, rather than on whether or not he acted with free will.

II. IMPLICATIONS OF FRANKFURT'S ACCOUNT
Following Frankfurt's lead, we can take the basic notion of freedom as the absence of obstacles to what one wants. The expressions free from and free to can then be seen as two branches of the same basic notion. Locutions involving the expression free from specify which obstacles are absent, and locutions involving the expression free to indicate the range of things available to the agent to do without obstacle. What Frankfurt and others call "freedom of action" is the absence of obstacles to doing what one wants to do; freedom of will is the absence of obstacles to willing what one wants to will

As Frankfurt's work makes clear, obstacles to doing what one wants to do can arise in two ways. They may have their origin in something external to the agent, such as social institutions, or in something internal to the agent, such as psychoses. It is easy to show that there can also be external and internal obstacles to freedom of will by adapting the familiar example of a man who chooses to stay in a room, unaware that the door is locked.⁵ Suppose that the man's

³ I am grateful to both Frankfurt and Alston for suggestions regarding these conditions.

Making precise the rather vague intuition behind this claim would take some doing and is beyond the scope of this paper. Perhaps it is enough for present purposes to say that by 'obstacle' here is surely meant something like "an obstacle which is, for all practical purposes, at the moment, physically insuperable."

Discussed the rather vague intuition behind this claim would take some doing and is beyond the scope of this paper. Perhaps it is enough for present which is, for all practical purposes, at the moment, physically insuperable."

Discussed in Frankfurt, "Alternate Possibilities and Moral Responsibility," op. cit.; cf. John Martin Fischer, "Responsibility and Control," this JOURNAL, LXXIX, 1 (January 1982): 24–40.

Digitized by Arya Samaj Foundation Chennai and eGangotri volition to stay in the room is produced in him by means of some device implanted in his brain and controlled by scientists who want him to stay in the room. In this case what the agent wills is not what he himself wants to will; he wills what the scientists want him to will. There is thus an external obstacle to the agent's willing what he wants to will, and so he does not act with free will in staying in the room.

Ordinarily when we ask whether a person acts with free will, what we are asking is in effect whether there is an external constraint of this sort on his will. The issue between compatibilists and libertarians, for example, can be understood at least in part as a dispute over whether the causal influences that compatibilists claim operate on a person constitute an external constraint on his will. Frankfurt's definition of freedom of will is a strong one, because, in order to have free will in his sense of the term, something more is required than the simple absence of external obstacles to willing what one wants. What else is needed can be seen by considering a revised version of the example of the man in the room. This time suppose that the man does want to leave the room—say, in order to get to his classroom to teach—and that the door is not locked, but that there is a black cat asleep on the lintel over the door and he is superstitious about black cats. He struggles with his superstitious fear, but finally his desire not to have to walk past the black cat gets the better of him; he gives up the struggle and wills to stay in the room. As before, he stays willingly, but he does not stay with freedom of will, in Frankfurt's sense of the term. In this case, however, the obstacle to his willing what he wants to will is internal rather than external, because it is his own desires which are the impediment to his freedom of will. Frankfurt's sense of 'freedom of will' is thus different from the ordinary sense of the term which we have in mind when we ask, for example, whether Aquinas believes in free will rather than theological determinism. It seems to me more nearly the sense of 'free' in the theological claim "You shall know the truth and the truth shall make you free." If it were not for the clumsy locution, we might call Frankfurt's sense 'complete freedom of will' since it encompasses and exceeds the ordinary sense of free will as absence of external obstacles to willing what one wants.6

There are, then, four basic sorts of obstacles to what one wants and four corresponding basic senses in which someone can be free in virtue of the absence of one or another of those obstacles: r

to

⁶ Frankfurt's account is usually employed by those who want to defend compatibilism, but both the revised account and Frankfurt's original views seem to me perfectly consistent also with incompatibilism.

Digitized by Arya Samaj Foundation Chennal and eGangotri

- (1) having no external obstacles to doing what one wants to do.
- (2) having no internal obstacles to doing what one wants to do.
- (3) having no external obstacles to willing what one wants to will, and
- (4) having no internal obstacles to willing what one wants to will.

When an agent lacks freedom in one of the first three senses, we generally do not hold him morally responsible for what he does or fails to do. We would not blame a student for failing to complete his studies if he could not attend the university for lack of funds or if psychosis rendered him incompetent to study. Similarly, we would not hold the man responsible for failing to try to leave the room if his willing to stay in the room were technologically induced by scientists. 7 But most of us would be inclined to blame the faculty member whose failure to meet his class stemmed from a superstitious fear of black cats, and the fact that we would is instructive. The division between freedom in the first three senses and freedom in the fourth sense is similar to Frankfurt's distinction between acting freely and acting with freedom of will. Someone who has freedom in the first three senses, even if he lacks it in the fourth sense, acts freely and so is responsible for his action. Only someone who also has freedom in the fourth sense, however, acts with freedom of will, in Frankfurt's strong sense of free will.

III. THE REVISED FRANKFURT ACCOUNT

On Frankfurt's analysis, the concept of a person is marked by a single hierarchical distinction, between an agent's first-order desires and volitions and his higher-order desires and volitions. But traditionally philosophers have analyzed a human person into three parts: desires or passions, will, and intellect. If we revise Frankfurt's view to take account of the role of intellect, I think we strengthen it without losing any of its explanatory power. Aquinas held that an agent wills to do some action p (or bring about some state of affairs q) only if the agent's intellect at the time of the action represents p (or q), under some description, as the good to be pursued.8 Here it is crucial to understand that, in this context, 'an agent's intellect' or 'an agent's reasoning' does not refer to something which is solely rational. By 'an

Aquinas's understanding of the relation of will to intellect is complicated and cannot be adequately discussed here. Basically, his view is that the intellect moves

⁷ There are, of course, complicated cases in which the agent's lack of freedom in one of these respects is ultimately attributable to the agent himself. The student might be impoverished because he lost all his money gambling, or his psychosis might be cause he lost all his money gambling, or his psychosis might be one he brought on himself as a result of his use of certain drugs known to cause psycholic downstance is a direct cause psychosis. In such cases, the agent's inability to do what he wants is a direct result of something the agent does want. To the extent to which the inability is tied to what the agent does want. To the extent to which the inability even if it to what the agent wants, the inability does not diminish moral responsibility even if it does diminish the agent's freedom at the time of his inability.

Aguings's the agent's freedom at the time of his inability.

Digitized by Arya Samaj Foundation Chennai and eGangotri

agent's intellect' I mean just the computing faculty of an agent. So understood, an agent's intellect may formulate a reason for an action in a manner that is hasty, thoughtless, ill-informed, invalid, or in any other way irrational.

Furthermore, it is important to understand that an agent's reason for an action may also be only implicit and not an explicit or conscious feature of his thought. In a recent paper, Robert Audi9 has argued cogently that x may count as the reason for an agent's action even when the agent has not consciously formulated some reason xas a reason for his action but would nonetheless give x as the reason for what he did if asked for an explanation. On this view, then, it is possible that an agent's intellect have gone through some process which contributes to a certain action on the agent's part, without the agent's being aware of that process as it is occurring. So to hold, as Aquinas does, that an agent wills to do some action p only if his intellect represents p as the good to be pursued does not entail that an agent does an action willingly only in case he first engages in a conscious process of reasoning about the action. 10 Aquinas's view requires only that some chain of reasoning (even if invalid and irrational reasoning) representing p as the good to be pursued would figure in the agent's own explanation of his action. In what follows, discussions of an agent's reasoning should be understood in light of these caveats: the reasoning in question need not be either rational or conscious.

In the spirit of Aquinas's view of intellect's direction of the will, we can make the following first revision of Frankfurt's account of the self. An agent has a second-order volition V2 to bring about some first-order volition V1 in himself only if the agent's intellect at the time of the willing represents V1, under some description, as the good to be pursued. A second-order volition, then, is a volition formed as a result of some reasoning (even when the reasoning is neither rational nor conscious) about one's first-order desires.

It will be helpful to make one other revision in Frankfurt's account. His use of the term 'second-order volition' is ambiguous between an agent's second-order desire that is effective in moving

the will by presenting it with an understanding of the good. In so moving the will, the intellect acts not as an efficient cause but rather as a final cause. The will is a natural inclination or appetite for the good; and the intellect moves the will, without coercion, by showing it what the good to be pursued is in a particular set of circumstances. But the will also moves the intellect directly, as an efficient cause—for example, by directing it to consider certain things and to neglect others. Cf., e.g., ST Ia, q.82, a.4.

⁹ "Acting for Reasons," *The Philosophical Review*, xcv (1986): 511–546.

¹⁰ I am grateful to Bruce Russell for raising this issue in correspondence.

him to make the corresponding first-order desire his will and a second-order desire that is not effective in that way. On Frankfurt's account, an unwilling addict who wants not to have the desire for heroin be his will but who nonetheless succumbs to his heroin addiction has a second-order volition for a desire not to take heroin. So, however, does the reformed addict who wants not to have the desire for heroin be his will and who has succeeded in that endeavor. When discussing first-order desires, Frankfurt identifies an agent's firstorder volition as a first-order desire which the agent makes his will and on which he acts. Consequently, although an agent may simultaneously have conflicting first-order desires, he cannot simultaneously have conflicting first-order volitions. For the sake of clarity in what follows, I want to make a second revision of Frankfurt's account by disambiguating the sense of 'second-order volition' along the lines of Frankfurt's distinction between first-order desires and first-order volitions.

If an effective desire is one which moves the agent all the way to action, then an effective second-order desire is one which moves the agent all the way to the action of making the corresponding firstorder desire his will. So a second-order desire constitutes a secondorder volition only if it is an effective desire and the agent has a first-order volition corresponding to it. On this usage, the reformed addict has a second-order volition not to have the desire for heroin be his will, but the unwilling addict who succumbs to his addiction does not. He has a second-order desire not to have the desire for heroin be his will; but because the second-order desire is not an effective desire, it does not constitute a second-order volition. To express Frankfurt's concept of freedom using this revised understanding of second-order desires and volitions, we should say that an individual has freedom of the will just in case he has second-order desires, his first-order volitions are not discordant with his secondorder desires, and he has the first-order volitions he has because of his second-order volitions.

As is the case with first-order volitions, it is not possible for an agent to have conflicting second-order volitions, but it is possible for him to have conflicting second-order desires. We might suppose, for example, that Verkhovensky in Dostoyevsky's The Possessed has both a second-order desire to have a desire for gambling, because the desire for gambling will make him well-liked by his friends, and a second-order desire not to have a desire to gamble, because stamping out the desire for gambling will win the admiration of Mrs. Stavrogin. It is worth noticing that where second-order desires conflict it will flict, it will not be possible for an agent to act on the corresponding first-order desires with freedom of will. Whether Verkhovensky has a

a

first-order volition to gamble or a first-order volition not to gamble, his will with regard to that volition is not free. In either case his first-order volition is opposed to one of his second-order desires, and so he fails to meet one of the criteria for freedom of will.¹¹

IV. CRITICISMS OF FRANKFURT & THE REVISED FRANKFURT ACCOUNT Of the many criticisms that have been raised against Frankfurt's account, three appear recurrently in the literature and seem especially worth considering. Various critics have charged that (1) on Frankfurt's account, an agent would be free with respect to a certain volition that was in accord with some higher-order volition even if that higher-order volition were directly produced somehow by someone else. But this is surely a counterintuitive result, because such an agent seems to be just as much the puppet of his manipulator as the agent whose first-order volitions are directly produced by someone else, and no one would want to say that the agent in the latter case has free will. ¹² Another recurrent complaint is that (2) Frankfurt's account of free will leaves us with an infinite regress of volitions. For an agent to act with free will with respect to some first-order volition

¹¹ The revised Frankfurt account has some affinities with Wright Neely's account of freedom, which holds that "an agent is free with respect to some action which he performed only if it is true that, if he had been given what he took to be good and sufficient reason for not doing what he did, he would not have done it" [Neely, "Freedom and Desire," The Philosophical Review, LXXXIII (1974):48]. On the revised Frankfurt account, an agent may have what he would agree, if asked, is good and sufficient reason for not doing some action x and yet still do x, because, under the sway of the passions (to take just one example), he interprets doing x at this time under some description which makes it mistakenly seem as if the good and sufficient reason does not apply to this particular action. But he would not be free with respect to x, even though he might be morally responsible for doing it, since in having what he takes to be good and sufficient reason against doing x, he also has a second-order desire not to do x. Neely has sometimes been interpreted as holding that what an agent really desires is what he desires when he is thinking rationally, i.e., without epistemic error. If this interpretation were correct (and I think it misses the force of the phrase 'what he took to be' in the preceding quotation from Neely's article), then the revised Frankfurt account would be opposed to Neely's, rather than similar to it, since on the revised Frankfurt account what an agent really desires can be based on a process of reasoning full of epistemic error. The revised Frankfurt account is also in some respects similar to the account of freedom given by Watson in his insightful paper "Free Agency," this JOURNAL, LXXII, 8 (April 24, 1975), 205, 290 Watson older Frenks 1975): 205–220. Watson objects to Frankfurt's account on the grounds that Frankfurt's furt has given us no reason to suppose that second-order desires represent what the agent himself wants just in virtue of being second-order. (I shall defend the revised Frankfurt account against this sort of objection in section IV below.) Watson substitutes an account based on a distinction between what an agent desires and what he values, and he takes an agent's values to be those principles and ends which the agent desires when he is being rational. Watson is right to introduce intellect into the hierarchical notion of the self, but I think it is a mistake to suppose that the valuings which play a role in free will are just the agent's rational principles and ends. In "Free Action and Free Will" (op. cit.), Watson expresses dissatisfaction with his earlier view for the same reason.

See, e.g., Shatz, op. cit., pp. 468/9.
CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

d

n

n

if

n

e

e

r

n

nt

ie

d er

nt

h in

a

es

's

es k-

y

k-

ie d

i-

e

e

0

ie d

V1 apparently requires that he have a higher-order volition V2 with which V1 is in accord. But then it seems that V2 itself must be freely willed in order for the agent to be acting with free will; for V2 to be freely willed, however, requires a higher-order volition V3 with which V2 is in accord, and V3 itself will require a higher-order volition V4, and so on. 13 Finally, some critics have objected to Frankfurt's account on the grounds that (3) it rests on an unwarranted notion of what counts as "the real you" and on a false theory of what counts as external or alien to a person. When we say that freedom is basically a matter of doing what one wants and that what one really wants is determined by considering one's higher-order volitions, we clearly presuppose that the first-order desires not in accord with the second-order volitions are not something the agent himself really wants, that these repudiated first-order desires are not part of the agent's self but somehow external to it. But such an understanding of the real self is at best controversial. Why should we identify an agent's self only with his higher-order volitions? Why should we suppose that these higher-order volitions represent what the agent really wants? Has psychology not made us aware that the darker sides of our nature, the repudiated or repressed firstorder desires, are just as much part of our selves as our higherorder desires and the first-order desires we approve of?14 The revised Frankfurt account is, I think, not vulnerable to any of these criticisms.

On the revised account, 15 an agent forms a second-order desire by reasoning (rationally or otherwise, consciously or not) about his first-order desires; and a second-order desire is a direct result of an agent's intellect representing a certain first-order desire as the good to be pursued. Given this connection between intellect and secondorder desires, an agent cannot be a passive bystander to his secondorder volitions. To be a second-order volition, a volition must be the result of reasoning on the agent's part. Even if it were coherent to suppose that one agent, say, Verkhovensky, could directly produce some reasoning in the mind of another, such as Stavrogin, that reasoning would not be Stavrogin's but rather Verkhovensky's (or at any rate a product of Verkhovensky's reasoning). If Verkhovensky con-

See, e.g., Thalberg, "Hierarchical Analyses of Unfree Action," Canadian Journal of Philosophy, VIII (1978): 211-225.

For Frankfurt's own response, see "Three Concepts of Free Action II," pp. 1/2

¹³ See, e.g., David Zimmerman, "Hierarchical Motivation and Freedom of the Will, Pacific Philosophical Quarterly, LXII (1981): 358ff.

tinuously produced thoughts in Stavrogin, then Stavrogin would have ceased to be a person and would instead be something like Verkhovensky's puppet. On the other hand, suppose Verkhovensky produced thoughts in Stavrogin's mind only occasionally, so that Stavrogin remained a person. In the computations leading to an action, Stavrogin's own intellect would take cognizance of the thought Verkhovensky had produced in Stavrogin's mind, and Stavrogin would then either accept or reject Verkhovensky's thought as a result of Stavrogin's own reasoning (however tacit or irrational that reasoning may be). As Stavrogin acts, then, the first-order volition stemming from his reasoning and the accompanying second-order desire will be Stavrogin's, not Verkhovensky's. Either way, Stavrogin would not have any second-order volitions produced by Verkhovensky. So, on the revised Frankfurt account, an agent's second-order volitions cannot be produced by someone else.

As for objection (2), Frankfurt himself believes that there is no theoretical limit to the levels of higher-order desires which a person may have, and that in general only common sense and fatigue keep a person from entertaining ever higher levels of higher-order desires. 16 On the revised Frankfurt account, however, the claim that the levels of higher-order desires may be infinite does not hold. In formulating a second-order volition an agent is bringing reason to bear on a state of his will and either approving or rejecting it. But in forming a third-order volition, the agent is not reiterating the process gone through to formulate a second-order volition. On the model the revised Frankfurt account gives of forming second-order volitions, forming a third-order volition consists in reasoning about and either accepting or rejecting a second-order volition. So an agent has a third-order volition V3 to bring about some secondorder volition V2 in himself only if his intellect at the time of the willing represents V2, under some description, as the good to be pursued. But since V2 is a desire for a first-order volition V1 generated by reason's representing V1 (at that time) as the good to be pursued, V3 will consist just in reaffirming the original reasoning about V1 which led to V2. In forming a third-order volition and considering whether he wants to have the relevant second-order volition, the agent will consider whether a desire for a desire for

¹⁶ See "Freedom of the Will and the Concept of a Person," op. cit., p. 16. Frankfurt returns to this problem in his forthcoming paper, "Identification and Wholeheartedness," where he analyzes the notion of wholeheartedness and uses it as the basis for a response to this problem.

some action p (or state of affairs q) is the good to be pursued. But a desire for a desire for p (or q) will be a good to be pursued just in case the desire for p (or q) is a good to be pursued, and that in turn will depend on whether the agent considers p (or q), under some description, at that time, a good to be pursued. So a third-order volition that supports a currently held second-order volition is in effect just the expression of a reevaluating and affirming of the reasoning that originally led to V1. And, in the same way, a third-order volition that rejects a currently held second-order volition will just be an expression of the reevaluation and rejection of the reasoning that led to the second-order volition. A third-order volition, then, is a result of a recalculation of the reasoning that originally underlay a second-order volition.

I do not mean to suggest that third-order volitions must always collapse into second-order volitions. It is possible for an agent to have third-order desires that are not only distinct from, but even discordant with, his second-order desires. This is so because volitions and desires, like emotions, are not always immediately responsive to reasoning. Even after we are quite sure that a danger is entirely past, we may nonetheless continue to feel some fear; and for a while, until the emotion subsides, we may need to remind ourselves recurrently that there is no cause for fear. Similarly, a second-order desire may take time to fade even when the agent has repudiated the reasoning that generated it. Consider someone whose childhood among Southern Baptists has left him with a desire to avoid alcohol and a second-order desire to have such a desire. Suppose that this person subsequently repudiates his childhood religion and, among other things, joins a sailing club where beer is regularly served after races. He will then have a first-order desire to drink beer, in order to fit in and be companionable and perhaps also to indulge a newly acquired taste for beer; but he may also notice in himself, as a surviving trace of childhood inclination, guilt at drinking and a second-order desire to have the sort of will which wills not to drink. Then he will remind himself that he has repudiated his Baptist background and that there is no harm and even some positive good in drinking a few beers. In this case, then, he will have a third-order desire not to have the second-order desire bequeathed him by his upbringing. His third-

¹⁷ It is, of course, also possible that, in a case of the sort I have been describing, an agent might reevaluate his original reasoning, reject it as unsatisfactory, and yet adopt the same second-order volition as before, although for different reasons. In such a case, the third-order volition is a result of reevaluating the reasoning and reaffirming not the original reasoning but rather just its conclusion.

order desire is thus a temporary measure to bring his second-order desire into line with a change in his reasoning, when he has repudiated his former reasoning and found his second-order desires slow in adjusting to the change.¹⁸

Apart from such cases, however, a third-order (or any higher order) desire or volition will collapse into a second-order desire or volition. Even in the case of the beer-drinking lapsed Baptist, the third-order desire will always be just a reflection of a certain secondorder desire. The lapsed Baptist's third-order desire to have a second-order volition that wills to will drinking is just a result of his reasoning that (at the time of the willing, under some description) drinking is a good to be pursued, that a desire for drinking is consequently also a good to be pursued, and that therefore his secondorder desire for a will that wills not to drink is a desire to be repudiated. If this Baptist now returns to his childhood convictions and again becomes convinced that drinking is wrong, he will not now form a fourth-order volition not to have his previous third-order volition. Instead, he will have once more reevaluated the reasoning behind the desire to drink and this time rejected it. He will consequently change his second-order volition to accommodate this new alteration in his reasoning, but he will not form a fourth-order volition. The third-order volition stemmed from his efforts to repress the second-order desire habitual from his youth, and his change of beliefs will carry with it the cessation of those efforts. Any attempt, then, to describe his state in terms of a fourth-order (or even higher-order) volition will collapse into the formulation of a secondorder volition. So, on the revised Frankfurt account, the number of levels of higher-order desires is not infinite but is rather limited to two or three.

¹⁸ There are also cases in which an agent's reasoning is confused and warrants conflicting second-order desires. An agent who notices such a conflict in his secreasoning and form a third-order volition in consequence. Nothing in this account entails that the agent's sorting out of his reasoning and accepting only one side of the Possessed, Verkhovensky swings from being contemptuous of Mrs. Stavrogin and rejecting the desires designed to please her to being infatuated with her and Frankfurt account and Dostoevsky's portrayal of the vacillating Verkhovensky give a wijd explanation of the line "a double-minded man is unstable in all his ways." We desires, whatever his corresponding first-order volition is, it will be discordant with one of his second-order desires.

Freedom of the will on the revised Frankfurt account should consequently not be taken as entailing that the will is free only if an act of willing is in accord with some higher-order volition. On this view, the will is free with respect to a volition V just in case V is accepted by the agent because his intellect approves of V (at that time, under some description) as the good to be pursued, and there is no higher-order desire of the agent's with which V is discordant. 19 Second-order desires can fit this definition of freedom without postulating higherorder desires over them. A second-order desire is itself an expression of the agent's reasoning and therefore eo ipso accepted by the agent as approved by his reasoning.20 If an agent's reasoning approves a desire as the good to be pursued, it must also in the very same process approve the desire for that desire. An agent who at one and the same time unambiguously considered a certain desire as a good to be pursued and also rejected the desire for that desire as not good would not be sane. So, on the revised Frankfurt account, a secondorder volition may be a free volition without itself being the object of some higher-order volition.

Objection (3) has been given its most forceful presentation by Irving Thalberg. Why, he asks, should we identify ourselves with our higher-order volitions? Have psychologists (and, in particular, Freud) not shown us that the "darker, savage, and nonrational aspects [of ourselves] are equally—if not more—important"? (ibid., p. 224). And if an agent is to be identified with certain of his darker first-order desires, then, in making such desires his will, he would be willing as he himself wants to will. Consequently, Frankfurt is wrong in holding that an agent's freedom depends on his having second-order volitions and governing his first-order desires so that they are not discordant with those second-order volitions. An agent may be doing just what he himself really wants, and so be free, when he acts against his second-order volitions and follows his savage or non-rational desires.

¹⁹ From this definition of free will together with the description of the relation between intellect and will sketched above, it is easy to see that we can generate the three Frankfurt criteria for freedom of the will. When an agent's reasoning approves a first-order volition V1 as the good to be pursued at the time of willing, under some description, then the agent's will also forms a second-order desire for that first-order volition; and the intellect's (rational or irrational) approval of V1, manifested by the second-order desire, is at least part of what makes V1 a volition rather than an ineffective first-order desire.

Except for the special sort of case sketched in the example of the beer-drinking lapsed Baptist, where third-order desires fill this role.

"Hierarchical Analyses of Unfree Action," op. cit.

On the revised Frankfurt account, it will not be quite right to say that an agent is to be identified with his second-order volitions. An agent wills what he really wants and is thus free when his first-order volitions are not discordant with his second-order desires, not because the agent is simply declared to be more truly identified with his second-order than with his first-order desires, but rather because the agent's second-order desires are the expressions of his intellect's reflection on his will, and the agent is to be identified with his intellect. It is, of course, possible to recast Thalberg's criticism so that it is directed against the revised version of Frankfurt's account. Why, one could ask, should an agent identify himself with his reasoning faculty rather than with his first-order desires (or his emotions, or subconscious, or any other element of his nature)? But this version of objection (3) is based on a confusion; it depends for its plausibility on a failure to distinguish two different senses of identification.

To see that this is so, consider a fictional variation on the biblical story of Tamar and Amnon (2 Sam. 13:1-20). According to the biblical story, David's son Amnon fell in love with his half-sister Tamar, tried to seduce her, raped her, and then rejected her with hatred. Suppose (contrary to the story in Samuel) that Tamar became pregnant as a result of the rape and bore a son, who quickly grew to look just like his father Amnon, and that no acceptable provisions for the child's care would be available if Tamar rejected him. In such circumstances Tamar would no doubt be torn between conflicting attitudes toward the child. On the one hand, she will recognize that the child is not identical with his father but rather an independent person, who is entirely innocent of the crime that resulted in his conception, and in this spirit she will want to cherish the child and be a good mother to him. On the other hand, when she looks at the child, she will see in him the hated face of the man who raped her; and if she is an ordinary human being, feelings of hatred and revulsion toward the child will rise in her as she recognizes signs of the father in the son. Now, suppose that on one particularly bad day when the child comes running to her, instead of welcoming him she flares up at him for no reason and hits him, because on that occasion the revulsion toward him has gotten the upper hand in her.

As she attempts to sort out her thoughts and feelings after this event, two interpretations of her action are open to her. She could say to herself: "I love my son. How could I treat him in that way? I've lost control of myself." Or she could say: "I can't stand that Amnon-faced child; I hate his father, and the sad truth is that I hate him, too." Tamar has been divided between her first-order desire to

cherish the child and her first-order desire to reject the child; and as she reflects on what she has done, she is in effect asking herself with which half of her divided will she identifies herself. We might be inclined to say that she is not to be identified with either side of her conflicting first-order desires; rather, what she is is a person struggling with a divided will. Such an understanding of Tamar may be correct as regards her past state, but it ceases to be viable once she has attacked the child, because she must then decide how to react to what she has done; and which of the two interpretations Tamar places on her action will make a great difference to the way in which life goes on between her and her son. She will, for example, have a very different sort of relationship with the child if she identifies herself in her own mind as a hater of the child. In that case she is in effect assenting to her rejection of him, and she will consequently not repent but excuse her action to herself, thinking: "I couldn't help it; I hate him for what his father did to me."

So Thalberg is certainly right in holding that, on some occasions, an agent is in fact to be identified with her darker first-order desires, as in the fictional example of the unrepentant Tamar. Notice, however, that, even in such a case, the reason for postulating such an identification is a second-order volition on the agent's part. In identifying herself with her first-order desire to reject the child, Tamar is evidently assenting to that desire and thereby ceasing to assent to the first-order desire to cherish him. But to say she assents to the first-order desire to reject the child is to say that she has a second-order volition to have a first-order volition to reject the child. On the view of the relation between will and intellect sketched above, what has happened is that her reasoning faculty has (at that time) rationalized her rejecting the child and found it acceptable, and she has consequently formed a second-order volition to make her first-order desire to reject the child her will.

In this sense of identification, then, for an agent to identify herself with some part of herself, such as certain of her first-order desires, is for her to form a second-order volition that accepts or assents to that part of herself. On this sense of identification, it is clear that an agent may identify herself with any of her first-order desires, no matter how savage or irrational they may in fact be; and what an agent identifies herself with is clearly up to her and depends on her reason and will. But it is important to see that, contrary to what Thalberg supposes, this conclusion in no way undermines the hierarchical account of the self. For what distinguishes the Tamar who identifies herself with her darker desires from the Tamar who repents them is

not the presence of a first-order desire to reject the child (for that desire is present in both), but rather the presence in the unrepentant Tamar of a second-order volition assenting to those darker first-order desires. This second-order volition the repentant Tamar, who struggles against her desire to reject the child, clearly lacks.²²

But there is also another sense in which an agent can be identified with some part of her character. However erring and faulty it may be, Tamar's reasoning faculty is essential to her, as no other part of her character is. If she were incapable of emotion, or if she were to become apathetic through depression, it would still be possible to consider Tamar a person.²³ But if her reasoning faculty were destroyed, she could no longer be counted a person; certainly she could not enter into any personal relationships, and the ability to do so seems a hallmark of a person.

There is a second sense of identification, then, in which it is correct to say that any agent is always to be identified with her reasoning faculty (whether it functions well or badly). This is not the same as the sense of identification in which the unrepentant Tamar identifies with her first-order desire to reject the child. An agent does not have a reasoning faculty in virtue of some second-order volition assenting to it;24 and, in this second sense of identification, it is not up to an agent to decide what she identifies herself with. Rather, in this sense of identification, an agent is to be identified with what is essential to her as a person, namely, her reasoning faculty. So, on the revised Frankfurt account, we are not simply trading Frankfurt's assumption that agents are to be identified with their second-order volitions for the new assumption that agents are to be identified with their intellects. What we have done instead is show that an agent's reasoning faculty is integral to her existence as a person. If we look to know what she herself really wants, we must consider what her reasoning faculty (at a certain time, under a certain description) assents to.

²⁵ Even on the revised Frankfurt account, an apathetic Tamar would count as a person if she had second-order desires, since, without first-order desires, there cannot be in her any discord between first- and second-order desires. And even very apathetic or severely depressed people can have the second-order desire not to be depressed any more.

²² The same sort of analysis could be given of an agent who identified herself with her baser first-order desires without the sort of internal struggle I have postulated for Tamar.

²⁴ Although it is, of course, open to the agent to form a second-order volition regarding a first-order desire to destroy his reasoning faculty. We can imagine a to lobotomize himself through drugs or surgery in order to put a stop to the torments his reasoning always brings him.

(And, as the story of Tamar shows, what her reasoning faculty assents to may be something that is, objectively considered, quite irrational.)

Second-order desires represent an agent's reasoning since they stem from the reflection of an agent's intellect on her state of will. Therefore, an agent is to be identified with her second-order desires as much as with her reasoning; her second-order desires represent what her intellect assents to (and so what she assents to) among her first-order desires. Consequently, on the revised Frankfurt account, it is correct, contrary to objection (3), to hold that second-order desires represent what an agent really wants. In fact, this second sense of identification is presupposed by the first. We can agree with Thalberg that an agent such as the unrepentant Tamar identifies with her base first-order desires because her identifying with them consists in her forming a second-order volition assenting to them. But the reason why the presence of such a second-order volition suffices for supposing Tamar to be identified with her desire to reject the child is that the second-order volition stems from Tamar's reasoning faculty and, in virtue of that connection, indicates what Tamar herself really wants. So the sorts of considerations Thalberg raises against Frankfurt's account, when they are properly understood, not only do not undermine, but instead actually support, Frankfurt's hierarchical analysis of the self.

V. APPLICATIONS TO ISSUES IN THE PHILOSOPHY OF RELIGION The revised Frankfurt account of a person and of free will is very fruitful for understanding a variety of religious practices and doctrines, such as the practices of adult baptism, confirmation, asceticism, the doctrine of justification by faith, the nature of heaven and hell, and the point of Romans 7.25 But in the remainder of this paper I shall concentrate on the application of that account to just two Christian doctrines, namely, that God sanctifies some people and that he hardens the hearts of others.

The doctrine of sanctification includes as a central component the claim that God intervenes in the minds of some people in order to make them morally better than they would otherwise be. Many Christian theologians, however, also hold that human beings have free will, where 'free will' is to be understood in an incompatibilist sense. From these two views a number of puzzles arise. First, to be morally good one must freely will some moral good. But then it seems that it is not possible even for an omnipotent God to make anyone morally good, since it is not possible for anyone to cause an

²⁵ I pursue some of these topics in forthcoming papers.

individual freely to will anything (where freedom of will is understood in an incompatibilist sense). Second, even if it were possible for God to make an agent freely will a good he would not have willed otherwise, it does not seem as if his doing so could count as making that agent morally better. If God causes the agent to will some moral good, then we might attribute some moral goodness to God in consequence, but why would we attribute moral goodness to the agent, who is nothing but a puppet of God's will? Finally, if God could in fact make a person morally good, why would he not do so for all persons? How could a good God fail to impart such a benefit to all human beings, so that there would never be any moral evil on earth and no one would ever be brought to hell?

To see the appropriate resolution of these puzzles it is important to understand that, on the Christian doctrine of sanctification, those whom God sanctifies are Christians who are still struggling with moral evil in themselves. Besides holding the traditional Christian beliefs about God, such a person will also believe both that certain things (such as beating one's wife, for example) are wrong and should not be done and that he himself is engaged in some of these morally wrong practices. Consider, for instance, some Christian Patricius who beats his wife Monica.26 On the revised Frankfurt account, we will say that, because Patricius believes it is wrong for him to beat Monica, he forms a second-order desire to make the firstorder desire not to beat his wife his will. But when the fit of wrath is on him, he acts on his first-order desire to beat her. When the fit has passed, he laments his action and recognizes that by his own lights he should have acted on his general prohibition to himself not to beat her. Patricius does not have control of himself; he does not have the strength of will to make his first-order desires conform to his second-order desires, and he is not able to make himself have the will he wants to have.

Suppose that Patricius also recognizes that this is his state and prays to God for help. Patricius has reasoned that his beating Monica is an evil but that he is a failure at his efforts to stop it, and that he needs God's help to be the sort of man he himself can approve of. Patricius's prayer for help expresses a second-order volition for God to alter Patricius's first-order will. What *Patricius* wants is for God to change his will in such a way that he no longer wills to beat his wife. If

²⁶ Of course, the real wife-beating Patricius was pagan, according to the account left us by his son Augustine; and no doubt when the patience of his wife Monica had with his paganism.

God were so to alter Patricius's will, Patricius's first-order volitions would be in accord with his second-order desires; and, on the revised Frankfurt account, Patricius's will would consequently be free. In giving Patricius a first-order volition not to beat his wife, then, God would not be destroying Patricius's freedom of the will but actually establishing it, since while Patricius's first-order volitions are discordant with his second-order desires, he does not have free will, however free his action of wife-beating may be.

Of course, the strength of second-order desires may vary. As Augustine tells the story, when he prayed to God to give him a will for sexual continence, he made the mental reservation "But not yet." How exactly to characterize Augustine's second-order desire in this case is not certain; but it is clear that, if God had given Augustine a will for sexual continence on that occasion in response to such a prayer, he would have been acting against Augustine's own secondorder desires. The result would have been not to evoke or enhance Augustine's free will but to undermine it, because the consequent first-order volition for sexual continence would have been against Augustine's second-order desire to have continence "but not yet." So, in general, sanctification will be a slow process. In response to such half-hearted prayers for help as Augustine's, God can produce some alteration in an individual's first-order will, by strengthening, to the degree warranted by the prayer for help, those first-order desires which are in accord with that individual's second-order desires. But if he is not to destroy freedom of the will, God will not be able to produce a first-order volition unless the second-order desire in the prayer for help is like the whole-hearted turning of the will experienced by Paul on the road to Damascus. Even with a secondorder desire for God to alter his will, then, Patricius may find that his struggle against the habit of wife-beating takes some time to win.

On this understanding of sanctification, it is clear that God is not violating an individual's free will in sanctifying him, even when free will is understood in an incompatibilist sense. What God is doing in sanctification is altering an agent's first-order desires to bring them into accord with that agent's own second-order desires, so that, in sanctifying an agent, God is producing or enhancing the agent's freedom of will. Furthermore, it is also clear that the alteration of will God effects in sanctification really does produce moral goodness on the part of the agent. In being sanctified, the agent does not become God's puppet, a simple adjunct to God's will; on the contrary, in sanctifying him, God is helping that agent to have the will the agent himself wants to have. The consequent moral goodness has

its origin in the agent's own volitions, not just in God's, as the objection to the doctrine of sanctification had supposed. Finally, on this view of sanctification, it is clear why it is not possible for God simply to sanctify everyone. The process of God's sanctifying a person consists in God's bringing an agent's first-order desires into line with his second-order desires in response to the agent's second-order volition that God do so. Where the requisite second-order desires and volitions are absent, God cannot alter the first-order desires without undermining or destroying freedom of will.²⁷ And, as the objections to the doctrine of sanctification indicate, it is not possible for an alteration of an agent's will which undermines the will's freedom to result in moral goodness on that agent's part. God cannot make human beings morally better unless they will that he do so.

Someone might think that such an account of sanctification is guilty of Pelagianism. But it does not entail the claim that an agent can achieve sanctification primarily by the exertions of his own will. Instead, this account holds that all the work resulting in moral improvement is done by God in response to the agent's recognition that he needs God's help and his willing to have it. (Of course, a person's willing of God's help is itself a response to God's action in that person's life; but the relation of God's action to such willing is part of the doctrine of justification, and what is at issue here is just the doctrine of sanctification—and not even the whole doctrine of sanctification, since faith, hope, and love are left out of account.) It is true that, on this account of sanctification, an act of free will on the agent's part is necessary for God's work of sanctification, but such a view was also held by Augustine and Aquinas, who are scarcely noted for their adherence to Pelagianism.

As for the doctrine that God sometimes hardens hearts, the classic text is in Exodus 7–14, where God is said to harden Pharaoh's heart so that Pharaoh does not weaken and allow the Israelites to leave Egypt, although the exodus of the Israelites from Egypt is the ultimate object of God's actions in the story, and God punishes Pharaoh for his resistance. That this is not a unique or isolated instance of God's hardening a heart is made explicit in Paul's epistle to the Romans, where Paul generalizes on the story of Pharaoh and concludes that "God hardens whom he will." This doctrine raises two

²⁷ If an individual were a wanton, in Frankfurt's sense, it would be possible for God to alter his first-order volitions without undermining any freedom of his, because such an individual by definition has no second-order desires and so fails to meet one of the conditions for freedom of will. If there are any human beings who are wantons, they will then be exceptions to the claim I make here.

questions concerning God's goodness. First, how could a good God ever intervene in an agent's willing in such a way as to make that agent morally worse? Willing that a human being become morally worse seems the essence of malice, and bringing about a person's moral deterioration is the sort of action typically attributed to Satan. At any rate, it seems utterly incompatible with perfect goodness. Second, even if it were somehow compatible with God's goodness to make an agent morally worse, how could a good God punish that agent for the moral failing God himself has induced in him? On the contrary, since the will that initiated and produced the moral failing is God's, any punishment appropriate for that moral failing seems to be deserved by God, not by the agent manipulated by God. But, on the revised Frankfurt account, there is a way to understand the doctrine that God hardens hearts which satisfactorily answers these questions and which shows this doctrine to be the mirror image of the doctrine of sanctification.

At one point in his diaries, Goebbels pauses to reflect on his own reviewing of German newsreel footage that shows the incredible devastation of Poland wrought by the German armies. Insofar as Goebbels has any morals at all, he seems to hold a primitive sort of divine-command ethics with Hitler fulfilling the role usually assigned to God. So he has no moral compunctions about Poland's devastation; on the contrary, his reason seems to regard it as morally approvable in virtue of having been commanded by Hitler. In addition, Goebbels is determined to further the German war effort in any way he can; and if the war requires inflicting more suffering of the sort visited on Poland, Goebbels is calmly resolute in his willingness to inflict it. In short, Goebbels has first-order desires to wreak the sort of devastation Poland suffered whenever doing so serves Germany's interests, and he also has second-order desires accepting and assenting to those first-order desires.

Nonetheless, as he reviews the newsreel from Poland, those first-order desires begin to slip. It may be that successful efforts at repressing normal human compassion consume a great deal of psychic energy; and since Goebbels's complaint of fatigue is one of the most frequent themes of the diaries, we might suppose that sheer exhaustion is sapping his ability to make his first-order desires conform to his second-order ones. At any rate, Goebbels notices those first-order desires slipping, and he addresses an exhortation to himself: "be hard, my heart, be hard." This exhortation to himself expresses, in effect, a second-order desire. He wants his heart to be hard; that is, he wants to have a first-order will which assents to the sufferings of

Digitized by Arya Samaj Foundation Chennai and eGangotri the Poles and which is consequently in accord with his second-order desires. Patricius's prayer to God for help expresses a second-order volition that God govern Patricius's first-order will for him, and it shows that Patricius despairs of governing his first-order desires himself. Goebbels's exhortation to himself, on the other hand, indicates an attempt on Goebbels's part to reassert his own control over his first-order desires, and it expresses a second-order desire that Goebbels's first-order desires be in conformity with his second-order desires to make his will serve the German war effort.

Now suppose that because of fatigue or some similar reason Goebbels is unable to reassert control over his first-order desires. Then his first-order desires to inflict suffering on the Poles will weaken, and some first-order desires to relieve their suffering may even appear. In his moment of weakness, he might even (for example) order a shipment of food sent to Poland. A first-order volition to relieve the Poles is clearly morally superior to a first-order volition to cause their suffering. So, if Goebbels weakens in this way, we might suppose that he will in consequence become morally better in some sense. And yet this supposed moral improvement will occur by a sort of accident, and it will certainly be against Goebbels's will, since what Goebbels himself really wants is represented by his second-order desires. In fact, insofar as his first-order desires to ravage the Poles weaken and the opposing first-order desires intensify, Goebbels's free will is undermined, because the new first-order desires are an obstacle to Goebbels's willing what he wants to will and thus constitute a hindrance to Goebbels's freedom of will. Suppose that God were to respond to Goebbels's exhortation to himself to be hard as if it were the atheist's analogue to a prayer, and suppose that God hardened Goebbels's heart for him, supplying the strength of will Goebbels in his fatigue was missing. Such an action on God's part would in no way violate Goebbels's free will. Rather, God would in that case be giving Goebbels what Goebbels himself wants. He would, in effect, be preserving Goebbels's free will from being undermined or destroyed, because in strengthening Goebbels's first-order desires for the ravaging of the Poles, God keeps Goebbels's first-order desires from being discordant with his second-order desires and so helps Goebbels to fulfill one of the requirements for freedom of will.

Finally, in hardening Goebbels, God would, I think, not be making Goebbels morally worse than he would have been without God's intervention. One might suppose that, if the sort of strengthening God grants Patricius makes him morally better, then analogously the sort of strengthening given Goebbels must make him morally worse.

But the two cases are disanalogous in a way which vitiates this inference. Consider Patricius when he fails to act in accordance with his second-order desire assenting to the first-order desire not to beat his wife. In this case as I have constructed it, Patricius is engaged in a struggle within himself; in beating his wife, he acts against what he himself believes to be good. If it is right that an erring conscience binds (and I think it is), then, even if wife-beating were not in itself a moral evil, Patricius would be made morally worse by beating his wife in virtue of putting into action what he believes is morally bad. For the case of Goebbels to be analogous in a way that would warrant saying that God's intervention makes Goebbels morally worse than he would have been without God's intervention, Goebbels's state as he acts to help Poland would have to be a mirror image of Patricius's state as he beats his wife: Goebbels would have to believe that what he was doing in helping Poland was a good thing, and he would thus have to be acting in accordance with his moral beliefs. If this were Goebbels's state, it seems clear that hardening his heart would make him morally worse than he would otherwise have been. It would also violate his free will on the revised Frankfurt account. Given both considerations, a good God would never harden anyone's heart in such a case.

But the actual case of Goebbels under consideration here is different. In fact, Goebbels believes that helping Poland is a bad thing; his intellect and his second-order desires are against doing so. Consequently, his state in helping Poland is not the mirror image of Patricius's but rather just the same: in helping Poland he would be putting into practice what he himself believes morally bad. So he does not become morally better by helping Poland in this frame of mind; and, if an erring conscience binds, then there is even a sense in which he becomes morally worse by doing so when he believes it wrong. At any rate, it is not true that, by hardening Goebbels's heart in these circumstances, God makes him morally worse by preventing an action that would contribute to Goebbels's moral improvement. In such a case, and only in such a case, God can harden a person's heart without making him morally worse than he would have been otherwise.

One might suppose that, even so, Goebbels is made worse just by having his evil first-order desires strengthened. If we suppose that Patricius is made morally better in virtue of having his good first-order desires strengthened even when they are not strengthened enough to constitute an effective desire, a volition on which he acts, then the mere fact that Goebbels's evil first-order desires are

strengthened makes him morally worse than he was before. I am willing to concede this claim. The problem with the radical sort of evil in which a moral monster such as Goebbels is sunk is that it blinds and distorts the conscience, 28 and even an erring conscience binds. Consequently, Goebbels is made morally worse than he was before whether he acts to afflict the Poles or refrains. If he refrains from destruction of the Poles, he undermines the German war effort in Poland and so betrays his chosen master, Hitler. He thus knowingly does what he believes to be wicked and is consequently made morally worse. On the other hand, since the destruction of Poland is in reality an objective evil, the desire for contributing to that destruction (apart from mitigating circumstances, absent here) is also evil: and the intensification of an evil desire does perhaps imply intensified evil in the desirer. If that is right, then, because Goebbels's conscience is distorted, his moral evil will increase whether God hardens his heart or leaves him alone.

The question, then, is which of the options open to God, hardening Goebbels or leaving him alone, produces the lesser of two evils; and there are several reasons for supposing the answer is hardening Goebbels's heart. As I have been at pains to show, by hardening Goebbels, God preserves a certain freedom of will for Goebbels which would be undermined if Goebbels acted against his moral beliefs. Second, it seems to me arguable that there would be some moral appropriateness in God's fulfilling Goebbels's quasi prayer for hardness of heart. There is something at least morally dubious about a villain such as Goebbels falling into some creditable action simply through fatigue. And, just as it seems right for God to answer Patricius's prayer for help by granting him the strength of will he wants to have, so it seems that Goebbels is getting what he deserves if in answer to his quasi prayer God grants him the strength to persist in the evil he has resolved on. Finally, it may be that in the case of a desperately evil man, such as Goebbels, giving him the strength to have as wicked a will as he wants is hazarding a last shot at reforming him. In hardening him, on my account, God acts only on Goebbels's first-order desires, strengthening them to bring them into conformity with his evil second-order desires; but Goebbels's intellect and second-order desires are in no way cemented in their evil when God hardens Goebbels's heart. In giving Goebbels the first-order desires

²⁸ What makes Goebbels a moral monster is that his intellect and second-order volitions are on the side of a major evil *and* he has no dissenting second-order desires. If he had dissenting second-order desires, then it would not be the case that God enhanced his freedom of will by hardening his heart.

he wishes, God may be providing Goebbels with a mirror, in his character and its consequences, to show him the evil of his wish; and Goebbels's understanding the evil of his second-order desire is the requisite first step to straightening his distorted conscience, to reforming his reason and second-order desires, and thus to beginning a moral rebirth.

In consequence, I think, we have our answers to the two questions raised above concerning the doctrine of hardening the heart. On this understanding of the doctrine, God hardens a heart when he strengthens evil first-order desires so that a second-order will bent on such evil may maintain its control over those first-order desires when fatigue or some other nonmoral accident might have caused the control to weaken. In doing so, God is not violating the agent's free will, and he is also not causing the agent to become morally worse than he would otherwise have been. And if God were to assign blame and punishment to an agent for what that agent did in consequence of God's hardening his heart, God would in no way be unjust to him, for the agent's own will is the source and origin of the evil the agent does. So, when God is said in Exodus to harden Pharaoh's heart, we should understand the text as claiming that God is doing for Pharaoh just what Pharaoh wants and lacks the strength to do for himself, namely, making Pharaoh's first-order volitions correspond to his evil second-order desire. In fact, this interpretation of the hardening of Pharaoh's heart helps to explain an otherwise perplexing feature of the story, namely, that God's hardening Pharaoh's heart alternates with Pharaoh's own hardening of his heart.

Someone might object to this account that it heartlessly neglects the welfare of the Poles, because whatever we might want to say about the merits of allowing Goebbels to fall into some momentary first-order desires to relieve Poland when we are considering only Goebbels's state of character, it is apparently undeniable that such a moment of weakness would have afforded the Poles relief. Therefore, just for the sake of the Poles, a good God would not harden Goebbels's heart. But this objection rests on the mistaken assumption that the fate of the Poles depends solely on the state of Goebbels's will. On the contrary, it is important to see that, if there is a God, there is no sort of inviolable connection between the state of Goebbels's will and the welfare of the Poles; and, from God's point of view, the fate of the Poles and the condition of Goebbels's will constitute two entirely separate issues. If Goebbels failed to have a moment of compassion for Poland's sufferings, it was still open to God to aid the Poles in some other way; and if in consequence of

God's hardening Goebbels's heart, Goebbels planned some new suffering for them, it would nonetheless be possible for God to interfere (in any number of ways) to prevent Goebbels from successfully accomplishing what he willed. If God does allow harm to Poland in consequence of some act of will on Goebbels's part, it is because of a separate decision on God's part about the Poles. What sort of decision that might be or how it might be justified are no doubt hard to explain, but no harder than any other instance of the problem of evil; and that problem lies just outside the scope of this paper.

In this way, then, the revised Frankfurt account, which gives a cogent and illuminating analysis of the nature of freedom and the concept of a person, shows divine sanctification and hardening of the heart to be mirror images of each other. In each case, God responds to an agent's desires by giving that agent the first-order volition he wants. When he hardens an agent's heart, he strengthens evil first-order desires in conformity to a second-order desire bent on that evil; and when he sanctifies an agent, he strengthens the first-order desires that the agent's second-order desires want as the good for that agent. In neither case does God's affecting an agent's first-order willing interfere with that agent's free will, on the revised Frankfurt account of freedom of will.

Frankfurt's basic idea of the will as commanding itself seems to me to have great explanatory power in more than one area of philosophy. That it should be particularly fruitful for philosophy of religion is perhaps not surprising given the central role of will in religion.

ELEONORE STUMP

Virginia Polytechnic Institute and State University

THE JOURNAL OF PHILOSOPHY

Digitized by Arya Samaj Foundation Chennai and eGangotri COMMENTS AND CRITICISM

FREE RIDING AND THE PRISONER'S DILEMMA*

N this note, I shall discuss Philip Pettit's "Free Riding and Foul Dealing." Pettit makes the important point that, among many-person dilemmas related to the provision of nonexcludable goods, one has to distinguish between the free-rider problem and the foul-dealer problem. He purports to show that this distinction can be made in terms of game-theoretic considerations, in terms of different many-person Prisoner's Dilemmas (PD). He also claims that this distinction has strategic significance related to the resolution of the dilemmas in question. Although I accept the importance of Pettit's distinction between the free-rider problem and the foul-dealer problem, I shall argue below that an exact distinction between these two cases cannot be made after all in terms of prisoner's dilemma games.

I.

To begin, consider Pettit's characterization of the free-rider problem. He chooses to define or characterize only what he calls the paradigm case. The paradigm free-rider problem is defined by the following four conditions:

- 1. There is a nonexcludable good attainable for a group: that is, a good enjoyed by all, though perhaps with different intensities, if enjoyed by any.
- 2. It can be attained, and rationally attained—whether at one of many levels, however smoothly related, or at the only level possible—by *K* members of the group, where *K* is less than all: the reward to each contributor in a subgroup *K* will exceed the cost of his contribution, though not necessarily by the same margin.

 It cannot be attained, or at least not rationally attained, by just one member of the group: the reward to the lone contributor will not cover the cost he has to bear.

4. The fear of contributing when the good is not produced, and the hope of not contributing when it is make it rational for each person not to contribute to the production of the good: specifically, it means that the strategy of not contributing maximizes expected utility (367/8).

Given the above characterization, Pettit defends several theses on free riderism. I shall consider the following three:

0022-362X/88/8508/0421\$00.70 © 1988 The Journal of Philosophy, Inc.

^{*} I wish to thank Philip Pettit and Martti Kuokkanen for comments and criticisms. This JOURNAL, LXXXIII, 7 (July 1986): 361–379.

(a) Every free-rider problem, so far as it is a prisoner's dilemma, is a type A dilemma. (A type A dilemma is an N-person prisoner's dilemma in which no cooperator is made worse off by a lone defector than he would become under universal defection. Otherwise, viz. if at least one cooperator is plunged below the universal defection line, the game poses a type B dilemma, which Pettit takes to define the fouldealer dilemma.)

(b) Under normal circumstances every type A dilemma constitutes a paradigm free-rider problem.

(c) The paradigm free-rider problem is usually (in most cases) a type A dilemma (366, 369)

I claim in this note that, whereas (a) is more or less trivially true, (b) and (c) are false. I also want to suggest that it may be a mistake to try to find neat game-theoretic explications for the free-rider and foul-dealer (and related) problems, although I do not want to deny the importance of game-theoretic considerations. In particular, I want to claim that Pettit's A and B type dilemmas are not as closely connected to the free-rider and foul-dealer problems as he claims. Let me now go into some detail.

II.

First, I would like to point out that there is clearly some free riding going on in the foul dealer's case as well, using the term 'free riding' in an ordinary loose sense. Thus, whereas I do not want to dispute the important intuitive distinction here, it must be pointed out that Pettit's choice of technical terminology is partly stipulative. Let us consider Pettit's best example of the foul-dealer problem. It is a Hobbesian kind of case of contributing toward universal peace. If all but one contribute (disarm), the lone defector (the only one remaining armed) will be in a powerful position. We may, however, question whether the K = N - 1 agents really have succeeded in providing a public good. Have they provided world peace? Maybe, if the foul dealer never has to resort to arms to secure his superiority. Be that as it may, my main point here is that the foul dealer has received something, his superior power, free from the contributors. So in a sense he clearly is a free rider. I agree with Pettit, though, that, at least in ideal or pure cases, it can be said that the free rider tries to benefit by the efforts of others, and the foul dealer to benefit at their expense. It seems to me that the notions of positive contribution and negative or harmful interference should be helpful in making the distinction. In the case of the free rider, there is lack of positive contribution (and thus omission) but no harmful interference. In the case of the foul dealer, there is in addition such harmful interference.2

² Cf. my A Theory of Social Action (Boston: Reidel, 1984), ch. 5.

Let us now consider the truth of thesis (a). It is worth noting that such dilemmas are really chicken games as far as the contributors are concerned, as for them universal defection (noncontribution) is the worst outcome. It is the noncontributors that make the whole *N*-person game a prisoner's dilemma. Let us now consider the central question here: Is every case of the paradigm free-rider problem which also is a PD an A type of case? Pettit says that this is so

because the lone free rider has exactly the effect of the lone defector in that sort of dilemma. As he is alone in his defection, N-1 others must have contributed to the good. But N-1 is the upper bound on K; so K or more must then have contributed. That means that the collective good must have been achieved and that everyone is better off than under universal defection. Thus the lone free rider does not plunge anyone below the baseline of universal defection (370).

The entailment in question holds basically because in clause 2 of the definition of the paradigm free rider there is the requirement that the contributors not lose by their contribution; and it is assumed here that the lone defector also gains relative to the universal defection line on the basis of clause 1. So the assumed PD will not allow anyone to go below the universal defection line. Given this, thesis (a) is obviously true. But, even so, I would like to make a couple of comments, which do not directly affect the truth of the thesis in question. One minor issue to be mentioned here relates to what is meant by having a free-rider problem at hand. Pettit assumes above that K agents in fact have contributed (and thus provided the good). This is an assumption that exceeds the above characterization of a paradigm free-rider problem, for clause 2 says only that the good can be provided by K agents, not that it actually has been provided. (But as we may understand also the PD in question analogously in terms of capability, this issue does not affect the truth of the thesis.) Another small problem is that Pettit concentrates too heavily on cases where there is one and only one defector (even if his definition allows for the presence of more than one defector). But obviously an instance of the free-rider problem typically has more than one defector (cf., e.g., tax paying). Analogously, a type A dilemma is not properly characterized by reference to one defector only, if real life applications are to be kept in mind.

Pettit says above that, in the free-rider case, everyone will be better off than under universal defection. This we can accept in the case of the contributors, to be sure. How about the lone defector? I have above accepted Pettit's claim that, as he also can enjoy the provided Public (or nonexcludable) good, he will be better off than under

universal defection. That is the whole point about free riding (and the intended reading of clause 1 of Pettit's definition), it may be argued. But, strictly speaking, clause 1 is compatible with the possibility that the costs of the free rider are so great (for example, loss of reputation or something of the sort) that he himself goes below the universal defection line, even if he is able to enjoy the public good. [This is a matter that does not as such affect the truth of Pettit's thesis (a).]

And to deal with a more realistic situation, consider the case of two or more defectors. It could happen in accordance with Pettit's definition of the paradigm free rider that one defector drives another defector below the universal defection line. I would also like to emphasize that even a sole defector can, in accordance with both Pettit's definition of the paradigm free rider and his definition of the type A dilemma, operate at the expense of the contributors, viz. by his defection cause the contributors to gain less than they otherwise would have gained (even if the defector will not, by definition, be able to drive any of the contributors below the universal defection line). This last remark indicates that the distinction between the paradigm free rider and the foul dealer is not as sharp as Pettit claims.

Consider next (b), which says that under normal conditions every type A dilemma constitutes a paradigm free-rider problem. Let us start by considering Pettit's construal of the notion of normal conditions. "Normal conditions obtain so long as no one can think that if he cooperates it is more probable that others cooperate, and sufficiently more probable to raise the expected value of cooperating above that of defecting" (371). Well, if we understand normal conditions in this way, then trivially defecting manages to maximize expected utility, and clause 4 of the definition of the free-rider dilemma is satisfied. But let me say that I see no good reason to take normal conditions in this extraordinarily strong sense, which I find partly question-begging. For surely normal conditions understood in this sense are often violated, as experimental research on PD shows, and as our ordinary experience seems to show.

How about the other clauses? They are false, I assert. Clause 1 is false, because nothing in the characterization of a PD excludes the possibility that it can deal with an excludable good. A PD is defined through the participants' rankings of outcomes, and those outcomes can well constitute (or be otherwise intimately connected with) an excludable good. Pettit says that "this good must be available to all, including the defector, since otherwise his deprivation would presumably motivate him to cooperate too" (370). But here Pettit makes

an extra motivational assumption which is not part and parcel of the definition of PD; and, whereas that definition says that the defector in this situation prefers defection to cooperation, it could be due to his achieving an excludable good in this case. Clause 2 is false if the standard definition of an N-person PD game is adopted, as then we do not require that there be any subgroup of the N participants which is able to secure the good in question. But, we may follow Pettit and define an N-person PD alternatively so that the provision of the collective good by means of $K (\leq N)$ contributors is guaranteed, and then the entailment trivially holds. (An analogous comment applies to clause 3.) So, Pettit's thesis (b) has been argued to be false on several grounds, of which the criticisms against clause 4 are weightiest.

As to thesis (c), it suffers from vagueness and is therefore somewhat hard to evaluate. I shall, however, assert that, given an understanding of free riding in a sense broader than Pettit's, there are proper examples of free riding which are not PD cases. In this sense, (c) is false. But, even if we construe free riders in Pettit's paradigmatic sense [and take (c) literally], I claim that there are instantiations of the free-rider problem which fail to be prisoner's dilemmas, strictly interpreted (see below). Then we are left with only the statistical problem of determining what usually is the case in the social world, and that is not a philosopher's task.

What reasons are there for believing that there are cases of paradigm free riders which are not cases of PD? First, there are considerations related to the number K of contributors (e.g., when an agent knows that K-1 have contributed) and the possibly lumpy nature of the good, which Pettit himself discusses but dismisses on the grounds of rarity (369-370). I assert, however, that, in many small-group situations with only a handful of participants, such cases are still important. Next, the definition of the paradigm free rider simply does not strictly guarantee the dominance of defection nor even the Pareto inferiority of universal defection—if probabilistic considerations are allowed to figure in characterizations of paradigm freeriding situations, as Pettit's talk in clause 4 about expected utility suggests. The use of the notion of expected utility there allows for cases where the probabilities in question can compensate both for the lack of Pareto inferiority in plain utilities and for the lack of strict dominance, as we shall see.

There are game-theoretic situations which fulfill Pettit's clauses of free riderism but which are not prisoner's dilemmas. For instance, we might have a chicken game where defection can plausibly be taken to maximize expected utility. For simplicity, let us concentrate on the

two-person case and consider the following symmetric two-person game:

$$C_1C_2 = \langle 3,3 \rangle$$
, $C_1D_2 = \langle 2,4 \rangle$, $D_1C_2 = \langle 4,2 \rangle$, $D_1D_2 = \langle 1,1 \rangle$

where 'C' and 'D' mean cooperation and defection, respectively. Here the subjective probabilities of the players might be such as to make defection maximize their expected utility. Recall that the game of chicken derives its name from a contest where two youngsters drive towards each other on a road on a collision course. It seems obvious that, if we go to the corresponding three-person game where three youngsters similarly approach the crossing point of three roads, each driving on his road, we analogously have a three-person game of chicken. In it defection can analogously be argued to maximize expected utilities in the case of the three youngsters.

Other examples of games occasionally instantiating free riding seem to be provided by various alternation and turn-taking games. Also, the battle of the sexes game can, it seems, have instantiations where defection maximizes expected utility.

In any case, I have now shown that the paradigmatic free-rider problem cannot be taken to be very closely connected to the prisoner's dilemma—not at least unless the connections are made stipulatively and to some extent artificially. Let me also emphasize that, if my criticism is correct, it follows that we are not left with a fully satisfactory explication of the foul-dealer problem. For, as indicated, if we (presystematically) equate foul dealing with acting directly at the expense of others and making some of them worse than they initially were, type B prisoner's dilemmas do not cover all those cases.

I have accepted Pettit's explications and worked within his framework. To end this note, let me sketch an alternative approach or research program. The idea is to characterize such different kinds of free riding as Pettit's paradigm case and his foul-dealer case without recourse to prisoner's dilemmas and even without an attempt to reduce the free-rider problem to the game-theoretic framework and still less to any specific kinds of games. Within this broader framework, one may also define a more liberal notion of the free-rider problem. My suggestion is to require liberally that in a free-rider predicament there be a socially preferred outcome, viz. a cooperative outcome in some sense giving a greater joint gain to the partici-

³ In his footnote 8, Pettit refers to the possibility of relaxing condition 4 of his definition of the paradigm free-rider problem. But it is important to note that adopting such a liberalization serves to drive free riderism and A type dilemmas farther from each other than Pettit wants to have them; and this of course supports the gist of the present note.

pants than results from universal defection, but allows nevertheless that any participant can gain from lone defection, or a defection by relatively few participants. (This is meant as a partial definition only and I shall not attempt here to formulate a sufficient condition of free riding.⁴)

Obviously, there are several game-theoretic structures compatible with my above liberal partial characterization of the free-rider problem. In addition to PD, at least chicken, battle of the sexes, and typical turn-taking games can have free-rider instantiations. We may still relax the above characterization of the free-rider problem by allowing the second condition not to hold for every participant. Then various nonsymmetric games also qualify for characterizing the structure of the free-riding situation. Needless to say, much detailed work would be required to carry out this program for the *N*-person case. Leaving further discussion to another context, it can be concluded that, no matter how exactly one conceives of the free-rider problem, there are many other situations than those characterized by different kinds of the prisoner's dilemma game which qualify as bearers of the general free-riding dilemma.

RAIMO TUOMELA

University of Helsinki

METAPHOR WITHOUT MAINSPRINGS: A REJOINDER TO ELGIN AND SCHEFFLER*

atherine Z. Elgin and Israel Scheffler¹ object that I "dismiss extensionalist theories of metaphor on the ground that substitution of coextensive terms does not always preserve metaphorical truth" (331). This charge is surprising, not the least because the passage they quote next and the substitutivity argument to which

⁰⁰²²-362X/88/8508/0427**\$**01.30

© 1988 The Journal of Philosophy, Inc.

⁴ I would like to suggest that, to get a set of necessary and sufficient conditions, we add two more requirements to the mentioned ones. First, as pointed out in this note, we should allow explicitly that a free rider's defection may involve a cost to the other participants (especially the contributors). Secondly, as free riding can be regarded as a strongly action-related notion, we should add that a free rider has the specific conditional intention to defect, the condition being that K members contribute.

^{*} I wish to thank Ted Cohen, Jonathan Malino, and Avishai Margalit for comments on an earlier draft of this note.

[&]quot;Mainsprings of Metaphor" (henceforth: MOM), this JOURNAL, LXXXIV, 6 (June 1987): 331-335.

Digitized by Arya Samaj Foundation Chennai and eGangotri they refer make no such allegation. What I did write was that the substitutivity argument is "prima facie sufficient to rule out extensionalist theories of metaphor," to which I immediately added that extensionalists "will attempt to circumvent this difficulty by appeal to their notion of secondary extensions," a move which I then questioned because of general reservations and for one specific reason to which Elgin and Scheffler do not respond.² For reasons of space, I was unable to elaborate my objections to extensionalism in MAD but, rather than "failing to appreciate the firepower [secondary extensions] add to the extensionalist's arsenal" (MOM 332), I hinted that

-and how-they misfire. Elgin and Scheffler further claim that their own extensionalist account of metaphor, to which most of their note is devoted, is "stronger" (335) than mine. Because their presentation in MOM brings out certain features of their approach more explicitly than their previous discussions of metaphor, I would like to take this occasion to examine their account with the care it deserves.3 It is. furthermore, to the credit of extensionalists like Elgin and Scheffler that they give central place to metaphor in their theory of language, but the problems that arise with their explanation of metaphor are general problems with their extensionalism which are of interest in their own right. Both our approaches, finally, give central place to "context" in metaphorical interpretation. By exploring the differences between our respective ways of incorporating contextual features into a theory of metaphor, we can also clarify some general issues about the treatment of context in semantic theory.

Elgin and Scheffler give three reasons for the superiority of their theory over mine. (1) Because it is extensionalist and intension-free, it is the more "economical" of the two and, assuming it explains as much, it is therefore the more "powerful" (335). (2) In fact it does explain more, "and more that we particularly want to know, about metaphor" (*ibid.*). (3) It "makes no use of a distinction between linguistic knowledge and collateral information" (*ibid.*), a distinction whose absence "strikes [the authors] as a good thing" (*ibid.*). As for (1): admirable as *ontological* economy may be, it should be remembered that great savings of one kind often require significant costs of

² "Metaphor as Demonstrative" (henceforth: MAD), this JOURNAL, LXXXII, 12 (December 1985): 677–710; p. 684, fn 10, my emphasis added. My entire discussion of "extensionalism" in MAD was limited to this one footnote.

³ See Israel Scheffler, Beyond the Letter (Boston: Routledge & Kegan Paul, 1979); C. Z. Elgin, From Reference to Reference (Indianapolis: Hackett, 1983); and, the ancestor of them all, Nelson Goodman, Languages of Art (Indianapolis: Hackett, 1976).

another; it therefore remains to be shown that an extensionalist theory is not only more economical but also on balance *simpler* than an intensionalist one—and for this to be shown on empirical, explanatory grounds rather than simply assumed as a self-evident truth. Similarly for (3): it may or may not be a "good thing" to distinguish between linguistic knowledge and collateral information. That should depend on the explanatory work the distinction performs in the theories in which it is utilized; there is nothing to "strike" one as a priori good or bad about the distinction itself. The force of both (1) and (3) ultimately rest, then, on (2): Elgin and Scheffler's claim that an extensionalist theory offers the better *explanation* of the phenomena of metaphor. Let us turn now to an example they offer in support of this claim.

Suppose that Romeo's utterance of

(a) Juliet is the sun.

is true, interpreted metaphorically, and that

- (b) The sun is the largest gaseous blob in the solar system. is also true, interpreted literally. Nonetheless,
 - (c) Juliet is the largest gaseous blob in the solar system.

may be false *even* under a metaphorical interpretation of the term 'the largest gaseous blob in the solar system'. Given our standard principles of semantic compositionality, why does the substitution of these literally coextensive terms fail to preserve the truth value of

their respective metaphorical interpretations?

Elgin and Scheffler's answer exploits Nelson Goodman's idea of the secondary extension of a term and his suggestion that "literal meaning . . . is a matter of primary and secondary extension" (332). They concede that the metaphorical extension of a term is not determined solely or simply by its literal primary extension, but they nonetheless insist that it is determined by nothing more than extensions—namely, by the term's literal secondary as well as primary extension. On their view, the principle that "the metaphorical depends on the literal" should be interpreted to mean that "metaphorical interpretation is a function of literal meaning" (332, my emphasis), where 'literal meaning' is then construed extensionally à la Goodman. From this "it follows . . . that coextensive terms with different secondary extensions bear different metaphorical inter-

⁴ Note that Goodman does not refer in the cited passage [in "On Likeness of Meaning," in *Problems and Projects* (Indianapolis: Hackett, 1972), pp. 221–230, p. 227] to literal meaning specifically; the same criterion would presumably hold for sameness of metaphorical meaning.

pretations" (*ibid.*): in particular, 'the largest gaseous blob in the solar system' cannot be substituted for 'the sun', salva metaphorical interpretation, because they differ in their secondary extensions—'Apollo's flaming chariot', e.g., is said to belong to that of the latter but not of the former. In effect, then, Elgin and Scheffler revise the principle of compositionality, making the (primary) extension of the metaphorical sentence depend, not only on the (primary) extensions of its parts, but also on their "meanings," i.e., their primary and secondary extensions. Only when two terms agree both in their primary and secondary extensions are they intersubstitutable. Thus, the invalidity of the inference from (a) to (c) is explained in purely extensionalist terms.

Of course, by this criterion, not only this but any substitution of one nonreplicating term for another is ruled out. As Goodman argued long ago, if we require secondary coextensiveness for sameness of meaning, it can be shown that no two nonreplicating words ever have the same meaning. Therefore, if metaphorical interpretation is a function of literal meaning, and no two nonreplicating words ever have the same literal meaning, no two nonreplicating words should bear the same metaphorical interpretation and be intersubstitutable.5 Some may find this consequence simply much too restrictive. Even if one does not, however, it calls out for an explanation. Is it some special feature of metaphor which requires literal Goodmanian synonymy for substitutivity? Or are metaphorical contexts no different in this respect from so-called intensional ones? Or do the authors really believe that the substitutivity of one term for another is never predictable semantically—in which case why is (primary) coextensiveness sufficient for ordinary extensional contexts?6

There is a further problem with Elgin and Scheffler's extensionalist solution which concerns the explanatory power of their notion of literal meaning, which, in turn, hinges on their characterization of a secondary extension. According to their identity conditions for literal meanings, *every* difference between the secondary extensions of two terms constitutes a difference between their literal meanings.⁷

⁵ Goodman, *ibid.*, suggests how one might weaken this requirement for synonymy by excluding certain compound schemata. None of these refinements affects my general point, however.

This paragraph has profited especially from discussion with Ted Cohen.

Why this is so is not obvious; not just any difference of literal meaning, in particular, of secondary extension, will always be relevant to the substitutability of terms preserving their metaphorical interpretation. It is unlikely, e.g., that 'Apollo's flaming chariot' enters into the metaphorical interpretation of 'the sun' in (a); what difference, then, should it make that it belongs to its secondary extension but not to that of 'the largest gaseous blob in the solar system'?

Under what conditions, then, does some description or picture fall in the secondary extension of a given term? Elgin and Scheffler state that the sun-descriptions that comprise the secondary extension of 'sun' need not themselves be true of the sun; the words and phrases in the class of sun-descriptions are simply those "readily recognized" as such.8 But what is it to be "readily recognized" as, say, a Ø? About this, we are told nothing save that we can readily recognize Øs "although we have no rule that governs [the] instantiation [of the term \emptyset]" (MOM 332). Now, whether we do or do not have such a rule, it seems clear that what is "readily recognized" as a \emptyset may vary with individual and occasion depending on "context"—on the beliefs, expectations, and symbol system of the recognizer. 10 Which beliefs, expectations, symbol system, etc. (for short: "beliefs")? Surely not all of them on each occasion but, for lack of a principled way of fixing a prior limit on the number and kind relevant in a given case, just as surely none can be arbitrarily excluded beforehand. So, given Elgin and Scheffler's criterion of ready recognition, the extension of (a token of) 'sun-description' on any given occasion would seem to be no more determined, or constrained, than the speaker's entire corpus of "beliefs" on that occasion. In other words, the class of sun-descriptions which constitutes the secondary extension of someone's inscription or utterance of 'sun' would seem to be coextensive with the class of words and phrases he would use to express his "sun-beliefs" on that occasion. 11 These "sun-beliefs" need not be limited even to those the speaker himself believes to be true of the sun. Some may be false, and be known by him to be false: witness Elgin and Scheffler's own example, 'Apollo's flaming chariot', a "readily recognized" sun-description relative to our common myth-

⁸ Some "readily recognized" \varnothing -descriptions (such as, stereotypical descriptions) may indeed be false of \varnothing and yet bear on individuating its literal meaning; but it is doubtful that this is true of all or even of most. Thus, notwithstanding the authors' assertion, I do not see why 'Apollo's flaming chariot' should bear on individuating the literal meaning of 'sun'. On the contrary, because of the intimate relation between "meaning" and the determination of extension, one would think that in general (although perhaps not always) the opposite would be the case.

Note that this assertion is given no argument, leaving it a complete mystery how we accomplish this far from trivial task. Whatever one's philosophical scruples about meanings, definitions, and analytic truths, speakers may yet know rules for fixing or determining the extensions of terms, whether or not they are generally able to state those rules and in whatever way they "have," or employ, them.

This theme, or versions of it, is to be found throughout Goodman's writings from "The Way the World Is" to Ways of Worldmaking (Indianapolis: Hackett, 1978).

By 'sun-beliefs' I mean "beliefs" that would be attributed to the speaker using content clauses in which the term 'sun' occurs. Here I ignore problems about beliefs for which there exist no actual inscriptions or utterances (tokens).

ological "knowledge" despite the fact that we all know it to be false. In general, in order for an expression \emptyset to count as a sun-description for a speaker, it ought to be sufficient if he believes that his audience will recognize it as such—whether or not he believes that it is true of the sun or he believes that his audience believes that. But, if this is all that is required for a term to be a sun-description, then with each change, addition, or contraction in the speaker's "beliefs" about the sun, the class of his sun-descriptions will change and, with it, the secondary extension of his use of 'sun'. And if, for every change in the secondary extension of a word, there is a change in its literal meaning, then with every change in the speaker's "sunbeliefs" there will be a change in the literal meaning of his uses of the term 'sun'. Thus, given Elgin and Scheffler's characterization of "literal meaning," everytime we learn, or simply believe we have learned, something new about the sun, the literal meaning we attach to the term 'sun' also changes.

None of this, to be sure, entails that every two replicas (tokens) of a word (type) will always have different literal meanings, but it also provides no constraints that prevent this possibility from obtaining, and it certainly looks like it will be the typical case. For many, this consequence will be objectionable enough. In fairness to Elgin and Scheffler, however, it is at one with their denial of any distinction between linguistic knowledge and collateral information, and it is also in the spirit of their own inscriptionalist preference for tokens rather than types. Yet, even from their point of view, it should be clear by now that the appeal to "literal meaning" does no explanatory work. For their condition that "metaphorical interpretation is a function of literal meaning"—i.e., of the literal primary and secondary extensions of the term—simply amounts to saying that the metaphorical interpretation of a term 'Ø' is a function of its literal primary extension and (for lack of any way to restrict the class of "beliefs" relevant to "ready recognition" of Ø-descriptions) everything the speaker "believes" about \varnothing . Attaching the label "literal meaning" to these last two conjuncts provides us with no more of an explanation of the substitution conditions for metaphorical interpretation than does the conjunction itself.12

¹² The extensions of 'sun-description' tokens, each determined by what the speaker readily recognizes as a sun-description on that occasion, may nonetheless be 'determinate,' as Elgin and Scheffler emphasize. My question concerns not the determinateness of their secondary extensions but their explanatory power. There is, furthermore, a question whether Elgin and Scheffler's implied dependence on notions like belief and recognition threatens their extensionalism. If so, they will give extensionalist analyses of belief and attitudinal sentences; see Scheffler, "An

Digitized by Arya Samaj Foundation Chennai and eGangotri
A similar problem arises for Elgin and Scheffler's notion of mention selection which they introduce in order to explain how tokens of one type (literally coextensive replicas) which bear different metaphorical interpretations in their respective contexts "differ semantically" (332). "In a mention-selective application," they say, "an expression refers, not to what it denotes, but to mentions thereof" (333); e.g., 'sun' mention-selects, not the sun, but other sun-descriptions and sun-pictures. Since different replicas may select different mentions in their respective contexts-e.g., tokens of 'sun' which occur in a work of Ptolemaic astronomy mention-select 'moving celestial body', while those which occur in a Copernican text mentionselect 'motionless celestial body'-Elgin and Scheffler propose that metaphorical applications of different replicas may also bear different metaphorical interpretations because the replicas select different mentions: the inscription of 'sun' that applies metaphorically to Juliet mention-selects 'life-sustaining' and 'beauteous', while the one that applies to Achilles selects 'life-threatening' and 'terrifying' (333).

Here also Elgin and Scheffler do not mean by the word 'mention' exclusively inscriptions that denote, or are true of, what the mention-selecting expression denotes or is true of. If they did, then even a token of 'sun' occurring in a work of Ptolemaic astronomy would not mention-select 'moving celestial body'. Despite what Ptolemaic astronomers reasonably believed at one time, they were wrong: the sun was never denoted by, or a member of the extension of, 'moving celestial body'. 13 Instead, what Elgin and Scheffler appear to include in the class of "mentions" from among which a given token \varnothing may make its selection is the whole range of expressions that fall in \varnothing 's secondary extension, expressions which, as we have seen, need not be true of the primary extension of \varnothing and do not generally stand in any particular semantic relation to it. But, since this is so, it is unclear why they call such differences in selected "associations" (333)—associations which are mainly a function of the speaker's "beliefs" and

There are, to be sure, any number of strategies by which one might defend the intuition that the Ptolemaic astronomers, despite their false beliefs, were nonetheless referred. less referring to the sun with these descriptions. I do not know, however, whether, or how, these can be made to fit with the rest of Elgin and Scheffler's inscriptionalist philosophy of language.

Inscriptional Approach to Indirect Quotation," Analysis, XIV, 4 (March 1954): 83-90; and now C. Z. Elgin, "Translucent Belief," this JOURNAL, LXXXII, 2 (February 1985), 74-91. ary 1985): 74-91. A full discussion of this issue requires more attention than present space allows, but I would argue (on lines similar to those in the text) that the cost of extensionalizing belief sentences is loss of semantic structure which, in turn, renders them impotent for explanatory purposes.

Digitized by Arya Samaj Foundation Chennai and eGangotri not of actual extension or truth—ways in which the tokens "differ

semantically" (332, my emphasis).

Furthermore, the sense in which Elgin and Scheffler's own account is extensional should also now be understood in a narrower sense than usual. It remains true, as they emphasize, that mention selection is formally an extensional relation in the sense they stipulate: its range is a determinate class of expressions (tokens) and two tokens are co-mention-selective iff they mention-select exactly the same tokens (333). Their theory is, however, not extensional in the sense that it is exclusively or primarily the extensions of the terms in question, and relations among those extensions, which do the explanatory work of their theory. Were they to spell out why some but not other terms are mention-selected on a given occasion, they would have to take into account not only extensional relations like denoting or coextensivity, but also beliefs and other psychological notions.14

Despite the fact that they twice assert that literally coextensive replicas "can, and often do, diverge in mention selection" (333), Elgin and Scheffler never, indeed, explain what accounts for different mention selections. All they have to say about this lacuna in their explanation is summed up in one word: Context (see, e.g., MOM 334). But this one-worder lends itself to two interpretations: Either it may mean that it is only by closely analyzing the context that we will arrive at an explanation, or it may mean that at this point-when we

14 The closing remarks of fn. 12 also apply here. It should also be noted that Elgin and Scheffler's claim that "exemplification . . . is thus extensional" (333) is not as simple a matter as they seem to suggest. As Goodman already recognized (Languages of Art, pp. 54-56), in the vulgar what we say is exemplified is a property. Apart from his general aversion to properties, however, Goodman argues that properties—at least on one view—are not finely enough individuated for purposes of exemplification, that only entities with the identity conditions of linguistic objects draw the requisite distinctions. At the same time, he also recognizes that exemplification cannot be simply of a predicate, or label, since labels are restricted to languages or to individual speakers. Thus, whereas Socrates surely exemplified rationality for Plato, Socrates did not exemplify the English label 'rational' for Plato. What is needed, then, is something finer than properties but not as fine as predicates of particular languages. Goodman's solution is to

construe 'exemplifies redness' as elliptical for 'exemplifies some label coextensive with "red" (55).

But this extensionalist solution—to this case of nonextensionality, as it should by now be clear—even he concedes is much too broad. Goodman's final words on this matter are instructive:

The answer is that the lines may be drawn with any degree of looseness or tightness. While "exemplifies rationality," taken by itself, says only "exemplifies some label coextensive with 'rational' ", the context usually tells us a good deal more about what label is in question (56).

Here the reference to 'context' serves, not as the start of an explanation, but as a convenient wastebasket for the extensionalist's unsolved problems; 'context' in this sense is where explanation stops and where a problem, once deposited, can be safely

find ourselves looking at context—explanation stops. I find both readings in MOM.

Sometimes the authors claim to have "explain[ed] both metaphorical reference and metaphorical likening" (MOM 335, my emphasis). But, if that is their aim, they have yet to meet it. For, whatever else one might ask of an explanation, explicitness is surely a minimal condition; but it is not one satisfied by their account, which leaves so much of the burden of explanation on the shoulders of a shrouded "context."

Elsewhere, on the other hand, Elgin and Scheffler emphasize that there are no "recipes for determining metaphorical meaning," only "heuristics," "cues and clues about which aspects of the context and background might be relevant" (331); that certain "contextual considerations are plainly insufficient to determine exactly which labels are jointly exemplified" (334) and, hence, metaphorically expressed. Contextual factors may "direct . . . our attention to a neighborhood in which a correct interpretation might be located" (334), but these directions are not determinative, systematic, fixed, or rule-like—unlike the explanations of (at least some) literal interpretations.

This negative prognosis may be true enough; I myself would endorse it for the specific interpretive modes Elgin and Scheffler discuss, e.g., exemplification. But, if this is their view, Elgin and Scheffler should not claim that their account "explains more" (335) than mine; our real disagreement is over whether it is nonetheless possible to give any kind of explanation of metaphor. That is, if this is Elgin and Scheffler's opinion, both of us agree that a context-sensitive mode of interpretation like exemplification does not furnish determinate, rule-like material for a semantic explanation of metaphor-in contrast, say, to the way in which denotation is determinate and rule-like (even if we "have" no rules; cf. MOM 332). For similar reasons, I argued in MAD that there are no "recipes," or algorithms, for determining the interpretation of a metaphor, i.e., its propositional content, whose determination (in a context) is a function of presuppositions subject to the very indeterminacy to which Elgin and Scheffler are sensitive. 15 What I also argued in MAD, however—and

¹⁵ Compare with this my own discussion of the role of exemplification which I locate in the pragmatics, rather than semantics, of metaphorical interpretation (MAD 699, fn. 27). In addition to the considerations mentioned in the text, I would argue that, precisely because exemplification is a mode of symbolization not specific to language, it belongs, not to a speaker's task-specific linguistic (semantic) competence, but to his ability to apply that competence in use. Furthermore, exemplification does not seem to underlie the interpretation of all metaphors; e.g., metaphorical interpretations of natural kind terms sometimes draw specifically on presuppositions about their stereotypical properties (see MAD 700). And because

where Elgin and Scheffler appear to disagree—is that there exists another level of metaphorical interpretation, *metaphorical character*, which consists of fixed and general, though context-sensitive, rules which, in turn, are of the same semantic kind as those which govern one kind of "literal" language, namely, demonstratives. It is at this level of interpretation, and only here, where I would argue that it is possible to uncover semantic explanations of metaphorical interpretation, i.e., explanations of metaphorical interpretations which consist of semantic knowledge specific to metaphor.

This semantic knowledge underlying metaphorical interpretation. like the formally analogous knowledge underlying the interpretation of demonstratives, relates the content, or truth condition, of the expression to a context. Elgin and Scheffler's attempt to summarize this knowledge with the three words of "advice: Look to context" (335) omits its most important ingredients, however: (1) that 'context' here means specifically presuppositions as opposed to elements of the actual circumstance, including actual extensions, and (2) that the context in question is the context of utterance/interpretation of the metaphor as distinguished from the context, or circumstance, in which we evaluate its truth. Because we are typically concerned with the truth of a metaphor in its context of utterance, it is necessary, so I argued in MAD, to distinguish its character and content or, analogously, the interpretive role of the context in determining what is said and its evaluative role in determining the actual truth value of what is said (MAD 707-710).16 I would add now that there also exists cognitive significance specific to the metaphorical mode by which the content of a metaphor is presented—information in addition to its content in any particular context which is carried specifically at the

stereotypical properties are sometimes false of the referents of the terms, they cannot be exemplified by them. Therefore, important as exemplification may be to some metaphorical interpretations, it cannot be essential to metaphor as such. Finally, it should be noted that Elgin and Scheffler's objection that my semantics "does not explain how a metaphor likens the literal and metaphorical referents of a term" (MOM 335) fails to mention that my account does explain this datum (at least as much as it can be explained)—and also through the apparatus of exemplification (or through Tversky's structurally similar notion of assymetrical similarity)—but in the pragmatics. Since I discuss this subject explicitly in fn. 27, Elgin and Scheffler's objection is especially odd. For want of a better explanation of their apparent ignorance of this part of my paper, I can only conjecture that, inasmuch as resemblance is neither necessary nor sufficient for replication, their replica of MAD does not include fn. 27.

¹⁶ Because of the conceptual confusions that ensue from failure to draw these distinctions, it is incorrect for Elgin and Scheffler to describe metaphorical character as "advice"—unless they would also so describe the character of an indexical like 'I'.

level of its context-oriented character; it is this character–istic information which, at least in part, is often felt to be lost in literal paraphrase and which is also central to understanding the specific cognitive contribution that metaphors make *as* metaphors to the explanation of human action.¹⁷

What is wrong, in other words, with Elgin and Scheffler's tidy formulation of my theory is that they indiscriminately lump together the importantly different ways in which "context" functions in determining the truth of a metaphor under one of its interpretations. It is this general problem which also underlies the specific difficulties discussed earlier. If one's theory of metaphor is not to be descriptively inadequate, it is agreed that the information that determines the various contents of different metaphors must include more than "lexical meaning"—construed either extensionally or intensionally; it must incorporate all kinds of contextual presuppositions, common knowledge, individual beliefs, and associations (see MAD 695-700). But, unless one then draws some distinction like character/content, it is impossible to provide an explanation which purports to be a semantic theory (e.g., which makes substantive use of notions like literal meaning in its explanans) of presumably semantic phenomena (cf. MOM 335)—which is exactly how Elgin and Scheffler appear to conceive of their own theory.18 Their failure to draw any such distinction results, for example, in the individuation of "literal meaning" by unmistakable extralinguistic beliefs; i.e., the result is neither semantic nor much of an explanation at all.

There are also independent reasons to distinguish character and content, or the different roles of context in metaphorical interpretation. Consider some of the different kinds of "metaphorical incompetence": the different ways in which we can fail to "get" a metaphor right. Sometimes we fully understand the content of the metaphor but, out of ignorance of the facts, do not know its truth value. Other times we know that an utterance is to be interpreted metaphorically—i.e., we know that it is to be assigned a metaphorical character—but, for lack of knowledge of its context of utterance, we fail to learn the relevant presuppositions that determine its content in that context. In yet a third set of circumstances, we may not know

¹⁷ I discuss the cognitive significance of metaphorical character further in "Knowledge of Metaphor and Knowledge by Metaphor," forthcoming.

This problem is not specific to Elgin and Scheffler's extensionalism; cf. MAD 697, fn. 26 for a similar criticism of a nonextensionalist theory of metaphor.

I owe this phrase to Ted Cohen; for further discussion, see his "Figurative Incompetence," MS., and my discussion of metaphorical disagreement in MAD 701. The four kinds of incompetence mentioned below are not exhaustive; Cohen assures me that there are always 613 ways to get things wrong.

whether a given utterance is to be interpreted metaphorically or literally; whether to assign it its standard ("literal") character or that of a metaphorical expression. And, finally, in a fourth class of cases, we may know that some utterance is not to be interpreted literally without yet knowing what it is to interpret an utterance metaphorically; say, we do not know that a metaphor depends for its interpretation on a contextual parameter (as opposed, say, to its literal meaning), or on a particular parameter different from that which determines whether its content is true. In this last case, in short, we have not yet mastered the function of metaphorical character.

Of these four cases, the incompetence exemplified by the first two is due to ignorance of a straightforward extralinguistic kind: lack of knowledge of the facts and of presuppositions that include, with much else, what Elgin and Scheffler would call "collateral information." The third case is an example of recognitional rather than strictly interpretive incompetence; it involves a failure to assign an utterance (token) its correct, or intended, linguistic description (type). Here the speaker may not lack the requisite knowledge of language but he fails to apply it correctly. Only in the last case do we have, I suggest, a clear example of *linguistic* incompetence; the lack of a kind of knowledge of language proper, similar to the linguistic incompetence of one who does not know that the content of (a token of) the indexical 'now' depends on its time of utterance—or (in contrast to metaphor) on its actual rather than a presupposed time.

A good, or even just adequate, theory of metaphor should have the conceptual resources to articulate and explain the differences among these cases. But, in order to do this, it is necessary to utilize the character/content distinction (or a similar one) which presupposes, in turn, a distinction between linguistic knowledge and extralinguistic, or collateral, information. Such a semantic theory may not answer what, "for students of metaphor, is a (perhaps the) crucial question" (MOM 335), but it does take some important first steps toward understanding what a speaker knows when he knows the interpretation of a metaphor. And to know the answer to this question strikes me as a good thing.

University of Chicago

JOSEF STERN

²⁰ The boundaries of this notion of linguistic knowledge are not, it should be noted, themselves known a priori; what falls in a speaker's knowledge of language, as opposed to his extralinguistic knowledge, will be determined by empirical inquiry into the nature of the human faculty for language. Hence, this notion will not furnish the epistemological stuff for a linguistic doctrine of a priori truth and should avoid important objections to the analytic/synthetic distinction.

Digitized by Arya Samaj Foundation Chennai and eGangotri $NOTES \ \ AND \ \ NEWS$

The editors are pleased to announce a new book series in philosophy entitled "Studies in Epistemology, Psychology, and Psychiatry," to be published by E. J. Brill under the general editorship of M. A. Notturno. The series will be devoted to the publication of recent philosophical works in these disciplines, and especially in the areas in which these disciplines intersect. Such works may be of contemporary or historical interest, and of theoretical or practical significance. Correspondence may be addressed to Elisabeth Erdman, E. J. Brill, P.O. Box 9000, 2300 PA Leiden, Holland.

The Sixteenth Hume Conference, sponsored by the British Society for the History of Philosophy, will be held at the University of Lancaster, August 21–24, 1989. The organizing committee is particularly interested in receiving papers which relate issues in Hume's thought to intellectual developments in Britain and Europe in the eighteenth century, or which relate Hume to other thinkers in history. Papers should be sent no later than October 31. Correspondence may be addressed to Dorothy Coleman, Phil. Dept., Bowdoin College, Brunswick, ME 04011.

The Stanford Humanities Center will be offering eight external fellowships for 1989–90 intended for postdoctoral scholars and teachers in the humanities, or those in other fields working on related projects, who would be interested in spending the academic year at Stanford. The fellowships are primarily intended to enable fellows to pursue their own research and writing; however, recipients are also expected to devote about one-sixth of their time to teaching or in some other way contributing to intellectual life at Stanford. The deadline for application is December 1, 1988. Application materials and further information regarding eligibility, stipends, and selection criteria may be obtained by writing Morton Sosna, Stanford Humanities Center, Mariposa House, Stanford Univ., Stanford, CA 94305-8630.

The Department of Philosophy of the Graduate Center of the City University of New York is pleased to announce the appointment of Stephen Schiffer to the doctoral faculty as Distinguished Professor, effective September. Jerry Fodor, currently Distinguished Professor, will continue as Adjunct Professor. Rogers Albritton will be Visiting Distinguished Professor in the Fall. In connection with a new logic program, Professors Melvin Fitting and Kenneth McAloon, and Distinguished Professor Rohit Parikh have accepted appointments to the doctoral faculty, effective September.

The Department of Philosophy at the University of Colorado at Boulder is pleased to announce the appointment of Christopher Shields as Assistant Professor, effective September.

Digitized by Arya Samaj Foundation Chennai and eGangotri
The Department of Philosophy at Northwestern University is pleased to
announce the appointment of Peter Kosso, effective September.

The Department of Philosophy at the University of Notre Dame is pleased to announce the appointments of Lynn Joy as Associate Professor of Philosophy and History, and Alasdair MacIntyre as the McMahon-Hank Chair of Philosophy, effective September.

NEW BOOKS

- ALFORD, C. FRED: Narcissism: Socrates, the Frankfurt School, and Psychoanalytic Theory. New Haven: Yale, 1988. x, 242 p. Cloth \$24.50.
- ANGELL, J. WILLIAM, and ROBERT HELM: Meaning and Value in Western Thought: A History of Ideas in Western Culture, vol. II: The Medieval and Modern Development. New York: University Presses, 1988. x, 495 p. Cloth \$36.00, paper \$22.50.
- BAKER, JOHN: Arguing for Equality. New York: Methuen, 1988. xv, 182 p. Cloth \$35.00, paper \$11.95.
- BAKER, MARK C.: Incorporation: A Theory of Grammatical Function Changing. Chicago: University Press, 1988. viii, 543 p. Cloth \$59.95, paper \$21.95.
- BAKKER, GERALD, and LEN CLARK: Explanation: An Introduction to the Philosophy of Science. Mountain View, CA: Mayfield, 1988. x, 212 p. Paper \$14.95.
- BARBER, BENJAMIN: The Conquest of Politics: Liberal Philosophy in Democratic Times. Princeton: University Press, 1988. 220 p. Cloth \$25.00.
- BENNINGTON, GEOFFREY: Lyotard: Writing the Event. New York: Columbia, 1988. ix, 189 p. Cloth \$27.50.
- BIGELOW, PAT: Kierkegaard and the Problem of Writing. Tallahassee: Florida State UP, 1987. xii, 227 p. Paper \$18.95.
- BLUESTONE, NATALIE HARRIS: Women and the Ideal Society: Plato's Republic and Modern Myths of Gender. Amherst: Massachusetts UP, 1987. x, 238 p. Cloth \$25.00, paper \$11.95.
- BRAINE, DAVID: The Reality of Time and the Existence of God: The Project of Proving God's Existence. New York: Oxford, 1988. xvi, 381 p. Cloth \$74.00.
- CALLAHAN, JOAN C.: Ethical Issues in Professional Life. New York: Oxford, 1988. xvii, 470 p. Cloth \$39.95, paper \$24.95.
- CAPUTO, JOHN D.: Radical Hermeneutics: Repetition, Deconstruction, and the Hermeneutic Project. Bloomington: Indiana UP, 1988. ix, 319 p. Cloth \$37.50, paper \$17.50.
- CARROLL, DAVID: Paraesthetics: Foucault, Lyotard, Derrida. New York: Methuen, 1988. xviii, 219 p. Cloth \$35.00, paper \$12.95.
- CHAKRAVARTI, UMA: The Social Dimensions of Early Buddhism. New York: Oxford, 1988. vii, 238 p. Cloth \$27.50.

- CLARKE, D. S. JR.: Principles of Semiotic. New York: Methuen, 1988. x, 164 p. Paper \$13.95.
- COLLINGRIDGE, DAVID: Criticism—Its Philosophical Structure. New York: University Press, 1987. vi, 172 p. Cloth \$23.50, paper \$11.50.
- crowe, MICHAEL J.: The Extraterrestrial Life Debate 1750-1900: The Idea of a Plural of Worlds from Kant to Lowell. New York: Cambridge, 1988. Paper \$22.95.
- CURLEY, EDWIN: Behind the Geometrical Method: A Reading of Spinoza's Ethics. Princeton: University Press, 1988. xxi, 175 p. Cloth \$25.00, paper \$9.95.
- ELASSER, WALTER M.: Reflections on a Theory of Organisms. Frelighsburg, Canada: Orbis, 1988. v, 160 p.
- FACIONE, PETER A.: The Student's Guide to Philosophy. Mountain View, CA: Mayfield, 1988. x, 162 p. Paper \$9.95.
- FEYERABEND, PAUL: Farewell to Reason. New York: Verso, 1988. 327 p. Cloth \$42.50, paper \$15.95.
- FLETCHER, JOSEPH: The Ethics of Genetic Control: Ending Reproductive Roulette. Buffalo: Prometheus, 1988. xxi, 218 p. Paper \$14.95.
- FLOCON, ALBERT, and ANDRE BARRE: Curvilinear Perspective: From Visual Space to the Constructed Image. Berkeley: California UP, 1988. xxii, 243 p. Cloth \$38.50.
- FREPPERT, LUCAN: The Basis of Morality according to William Occam. Chicago: Franciscan Herald, 1988. vii, 191 p. Paper \$8.95.
- GENOVE, JUDITH: Power, Gender, Values. Alberta, Canada: Academic, 1987. vii, 190 p.
- GEOGHEGAN, VINCENT: Utopianism and Marxism. New York: Methuen, 1988. 164 p. Paper \$10.95.
- GOULD, CAROL C.: Rethinking Democracy: Freedom and Social Cooperation in Politics, Economy, and Society. New York: Cambridge, 1988. x, 363 p. Cloth \$39.50.
- GOVIER, TRUDY: Problems in Argument Analysis and Evaluation. Providence: Foris, 1988. viii, 302 p. Paper \$29.90.
- GRACIA, JORGE J. E.: Individuality: An Essay on the Foundations of Metaphysics. Albany: SUNY, 1988. xx, 315 p. Cloth \$44.50, paper \$14.95.
- GREEN, RONALD M.: Religion and Moral Reason: A New Method for Comparative Study. New York: Oxford, 1988. xvi, 278 p. Cloth \$29.92, paper \$10.95.
- HAGER, FRITZ-PETER: Gott und das Böse im antiken Platonismus. Wurzburg: Konigshausen & Neumann, 1987. 165 p.
- HAHN, ROBERT: Kant's Newtonian Revolution in Philosophy: Journal of the History of Philosophy Monographs. Carbondale: Southern Illinois UP, 1988. x, 143 p. Paper \$16.95.
- HAMPEL, CYRIL: Foundations of Philosophy: A Theory of Language, Reality, and Method. New York: Vantage, 1988. xlv, 316 p. Cloth \$19.95.
- HATTIANGADI, J. N.: How is Language Possible? Philosophical Reflections on

the Evolution of Language and Knowledge. La Salle, IL: Open Court, 1987. xxi, 224 p. Cloth \$29.95, paper \$13.95.

HILBERT, DAVID R.: Color and Color Perception: A Study in Anthropocentric Realism. Chicago: University Press, 1988. ix, 146 p. Cloth \$24.95, paper \$11.95.

HINES, TERENCE: Pseudoscience and the Paranormal: A Critical Examination of the Evidence. Buffalo: Prometheus, 1988. xii, 372 p. Paper \$17.95.

HOLLIS, MARTIN: The Cunning of Reason. New York: Cambridge, 1988. viii, 222 p. Cloth \$39.50, paper \$12.95.

HOLTON, GERALD: Thematic Origins of Scientific Thought: Kepler to Einstein. Cambridge: Harvard, 1988. 499 p. Cloth \$25.00, paper \$12.95.

HONNER, JOHN: The Description of Nature: Niels Bohr and the Philosophy of Quantum Physics. New York: Oxford, 1988. viii, 235 p. Cloth \$55.00.

HOOKWAY, CHRISTOPHER: Quine: Language, Experience, and Reality. Stanford: University Press, 1988. xi, 227 p. Cloth \$35.00, paper \$11.95.

JACOBSON, NOLAN PLINY: The Heart of Buddhist Philosophy. Carbondale, IL: Southern Illinois UP, 1988. xiv, 189 p. Cloth \$16.95, paper \$9.95.

JOY, LYNN SUMIDA: Gassendi the Atomist: Advocate of History in an Age of Science. New York: Cambridge, 1988. xiii, 311 p. Cloth \$34.50.

KAELIN, E. F.: Heidegger's Being and Time: A Reading for Readers. Tallahassee: Florida State UP, 1988. ix, 358 p. Cloth \$29.95, paper \$15.00.

KEARNEY, RICHARD: Modern Movements in European Philosophy. New York: St. Martin's, 1988. 346 p. Paper \$16.95.

KREISER, LOTHAR, SIEGFRIED GOTTWALD, and WERNER STELZNER: Nichtlassische Logik: Eine Einführung. Berlin: Akademie-Verlag, 1987. xiii, 437 p. KRIPS, HENRY: The Metaphysics of Quantum Theory. New York: Oxford,

1988. 257 p. Cloth \$55.00.

KULTGEN, JOHN: Ethics and Professionalism. Philadelphia: Pennsylvania UP, 1988. xii, 394 p. Cloth \$39.95, paper \$19.95.

LACHS, JOHN: Mind and Philosophers. Champaign: Illinois UP, 1987. 244 p. Paper \$15.95.

LAUER, QUENTIN: G. K. Chesterton: Philosopher Without Portfolio. New York: Fordham, 1988. 191 p. Cloth \$22.50.

LEAR, JONATHAN: Aristotle: The Desire to Understand. New York: Cambridge, 1988. xi, 328 p. Cloth \$39.50, paper \$12.95.

LIPMAN, MATTHEW: Philosophy Goes to School. Philadelphia: Temple UP, 1988. ix, 228 p. Cloth \$34.95, paper \$14.95.

LIVINGSTON, PAISLEY: Literary Knowledge: Humanistic Inquiry and the Philosophy of Science. Ithaca: Cornell, 1988. ix, 276 p. Cloth \$29.95, paper \$9.95.

LÜDEKING, KARLHEINZ: Analytische Philosophie der Kunst. Frankfurt: Athenäum, 1988. 230 p.

LYCAN, WILLIAM G.: Judgement and Justification. New York: Cambridge, 1988. xiv, 236 p. Cloth \$39.50, paper \$12.95.

MAQUET, JACQUES: The Aesthetic Experience: An Anthropologist Looks at the Visual Arts. New Haven: Yale, 1988. xi, 272 p. Paper \$17.95.

- MARGOLIS, HOWARD: Patterns, Thinking, and Cognition: A Theory of Judgement. Chicago: University Press, 1988. xii, 332 p. Cloth \$45.00, paper \$15.95.
- MARSHALL, DAVID: The Surprising Effects of Sympathy: Marivaux, Diderot, Rousseau, and Mary Shelly. Chicago: University Press, 1988. x, 286 p. Cloth \$27.50.
- MAVRODES, GEORGE I.: Revelation in Religious Belief. Philadelphia: Temple, 1988. 161 p. Cloth \$19.95.
- MECHANIC, JANEVIVE JEAN: The Logic of Decision Making: An Introduction to Critical Thinking. New York: Lang, 1988. ix, 205 p. Cloth \$35.50.
- MELLOR, D. H.: The Warrant of Induction: An Inaugural Lecture. New York: Cambridge, 1988. 25 p. Paper \$3.95.
- MILLER, RICHARD W.: Fact and Method: Explanation, Confirmation, and Reality in the Natural and Social Sciences. Princeton: University Press, 1987. ix, 611 p. Cloth \$65.00, paper \$14.50.
- MUDIMBE, V. Y.: The Invention of Africa: Gnosis, Philosophy, and the Order of Knowledge. Bloomington: Indiana UP, 1988. xii, 241 p. Cloth \$39.95, paper \$14.95.
- NYIRI, J. C., and BARRY SMITH, eds.: Practical Knowledge: Outlines of a Theory of Traditions and Skills. New York: Routledge & Kegan Paul, 1988. vii, 213 p. Cloth \$62.50.
- OKRENT, MARK: Heidegger's Pragmatism: Understanding, Being, and the Critique of Metaphysics. Ithaca: Cornell, 1988. xiii, 302 p. Cloth \$29.95.
- OLSON, KENNETH RUSSELL: An Essay on Facts. Chicago: University Press, 1988. vii, 105 p. Cloth \$23.95, paper \$11.95.
- ONIANS, R. B.: The Origins of European Thought about the Body, the Mind, the Soul, the World, Time, and Fate. New York: Cambridge, 1988. xviii, 583 p. Cloth \$59.50, paper \$19.95.
- POITIER, RICHARD: The Renewal of Literature: Emersonian Reflections. New Haven: Yale, 1988. 244 p. Paper \$9.95.
- POLLARD, CARL, and IVAN A. SAG: Information-based Syntax and Semantics. Stanford: CSLI, 1988. x, 233 p. Cloth \$39.95, paper \$17.95.
- PUTNAM, HILARY: The Many Faces of Realism. LaSalle, IL: Open Court, 1987. 98 p. Cloth \$19.95, paper \$8.95.
- RÉE, JONATHAN: Philosophical Tales. New York: Methuen, 1988. 162 p. Paper \$12.95.
- ROSEN, STANLEY: The Quarrel between Philosophy and Poetry: Studies in Ancient Thought. New York: Routledge & Kegan Paul, 1988. xiii, 223 p. Cloth \$29.95.
- ROTENSTREICH, NATHAN: Order and Might. Albany: SUNY, 1988. 238 p. Cloth \$39.50, paper \$14.95.
- ROTHMAN, MILTON A.: A Physicist's Guide to Skepticism. Buffalo: Prometheus, 1988. 247 p. Cloth \$19.95.
- SAGOFF, MARK: The Economy of the Earth: Philosophy, Law, and the Environment. New York: Cambridge, 1988. x, 271 p. Cloth \$29.95.
- SCHMIDT, DENNIS J.: The Ubiquity of the Finite: Hegel, Heidegger, and the

Entitlements of Philosophy. Cambridge: MIT, 1988. xiii, 241 p. Cloth \$22.50.

SCHNÄDELBACH, HERBERT: Vernunft und Geschichte: Vorträge und Abhandlungen. Frankfurt: Suhrkamp, 1987. 290 p.

SCHWANAUER, FRANCIS: Philosophical Fact and Paradox. New York: Lang, 1987. 244 p. Cloth \$37.50.

SHEPPARD, ANNE: Aesthetics: An Introduction to the Philosophy of Art. New York: Oxford, 1987. 172 p. Paper \$7.95.

SOLOMON, ROBERT C.: Continental Philosophy since 1750: The Rise and Fall of the Self. New York: Oxford, 1988. viii, 214 p. Cloth \$29.95, paper \$9.95.

SOUTHWELL, SAMUEL B.: Kenneth Burke and Martin Heidegger: With a Note against Deconstructionism. Gainesville: Florida UP, 1987. ix, 156 p. Paper \$15.00.

STAROBINSKI, JEAN: Jean-Jacques Rousseau: Transparency and Obstruction. Chicago: University Press, 1988. xxxviii, 421 p. Cloth \$60.00, paper \$19.95.

STEEN, W. J. VAN DER, and P. J. THUNG: Faces of Medicine: A Philosophical Study. Boston: Kluwer, 1988. vii, 236 p. Cloth \$69.00.

STEINBERGER, PETER J.: Logic and Politics: Hegel's Philosophy of Right. New Haven: Yale, 1988. xiii, 254 p. Cloth \$24.00.

THOMAS, GEOFFREY: The Moral Philosophy of T. H. Green. New York: Oxford, 1988. xvii, 406 p. Cloth \$72.00.

TOOLEY, MICHAEL: Causation: A Realist Approach. New York: Oxford, 1988. xiii, 360 p. Cloth \$65.00.

WALTERS, GREGORY J.: Karl Jaspers and the Role of "Conversion" in the Nuclear Age. New York: University Presses, 1988. xxvii, 300 p. Cloth \$29.50, paper \$15.75.

WHITE, ALAN R.: Methods of Metaphysics. New York: Croom Helm, 1988. 203 p. Paper \$12.95.

WIEBE, PHILLIP H.: Theism in an Age of Science. New York: University Presses, 1988. ix, 178 p. Cloth \$22.50, paper \$11.75.

WINCH, PETER: Trying to Make Sense. New York: Blackwell, 1988. viii, 213 p. Cloth \$34.95.

WINFIELD, RICHARD DIEN: *The Just Economy*. New York: Routledge & Kegan Paul, 1988. 252 p. Cloth \$35.00.

WOLFF, ROBERT PAUL: Moneybags Must Be So Lucky: On the Literary Structure of Capital. Amherst: Massachusetts UP, 1988. 82 p. Cloth \$20.00, paper \$8.95.

WOOLHOUSE, R. S.: The Empiricists. New York: Oxford, 1988. 183 p. Cloth \$29.95, paper \$9.95.

ZWICKY, JAN: Wittgenstein Elegies. Alberta, Canada: Academic, 1987. 70 p.

NEW BOOKS: ANTHOLOGIES

ACKRILL, J. L., ed.: A New Aristotle Reader. Princeton: University Press, 1988. xiii, 580 p. Cloth \$50.00, paper \$13.95.

- BAIRD, ROBERT M., and STUART E. ROSENBAUM, eds.: Philosophy of Punishment. Buffalo: Prometheus, 1988. 151 p. Paper \$8.95.
- BECK, LEWIS WHITE, ed.: Kant Selections. New York: Macmillan, 1988. ix, 478 p. Paper \$8.00.
- BELL, RICHARD H., ed.: The Grammar of the Heart: New Essays in Moral Philosophy and Theology—Thinking with Kierkegaard and Wittgenstein. San Francisco: Harper & Row, 1988. xxviii, 29 p. Cloth \$24.95.
- BERLINGER, RUDOLF, et al., eds.: Perspektiven der Philosophie: Neues Jahrbuch 1987, Band 13. Amsterdam: Rodopi, 1987. xxxix, 475 p.
- BERNAUER, JAMES, and DAVID RASMUSSEN, eds.: The Final Foucault. Cambridge: MIT, 1988. ix, 168 p. Paper \$9.95.
- CANTRIL, ALBERT H., ed.: Psychology, Humanism and Scientific Inquiry: The Selected Essays of Hadley Cantril. New Brunswick, NJ: Transaction Books, 1988. xiv, 250 p. Cloth \$34.95.
- CULLEN, BERNARD, ed.: *Hegel Today*. Brookfield, VT: Gower, 1988. xi, 121 p. Cloth \$41.95.
- DENEEF, A. LEIGH, et al., *The Academic's Handbook*. Durham, NC: Duke, 1988. xii, 287 p. Cloth \$20.00, paper \$9.95.
- FABB, NIGEL, et al., ed.: The Linguistics of Writing: Arguments between Language and Literature. New York: Methuen, 1987. vi, 325 p. Cloth \$45.00, paper \$14.95.
- FEKETE, JOHN, ed.: Life After Modernism: Essays on Value and Culture. New York: St. Martin's, 1988. xix, 197 p. Cloth \$29.95, paper \$12.95.
- FETZER, JAMES H., ed.: Probability and Causality: Essays in Honor of Wesley C. Salmon. Boston: Reidel, 1987. xvii, 353 p. Cloth \$79.00.
- GARDENFORS, PETER, and NILS-ERIC SAHLIN, eds.: Decision, Probability, and Utility: Selected Readings. New York: Cambridge, 1988. x, 449 p. Cloth \$49.50, paper \$15.95.
- GAZDAR, GERALD, et al.: Natural Language Processing in the 1980s: A Bibliography. Chicago: University Press, 1988. vii, 240 p. Cloth \$29.95, paper \$11.95.
- GILMAN, SANDER L., ed.: Conversations with Nietzsche: A Life in the Words of his Contemporaries. New York: Oxford, 1987. xii, 276 p. Cloth \$26.00.
- GINSBERG, MATTHEW L., ed.: Readings in Nonmonotonic Reasoning. Los Altos, CA: Kaufmann, 1987. viii, 481 p. Paper \$26.95.
- GOTTHELF, ALLAN, and JAMES G. LENNOX, eds.: Philosophical Issues in Aristotle's Biology. New York: Cambridge, 1987. xiii, 462 p. Cloth \$49.50, paper \$15.95.
- GREELY, ROGER E., ed.: The Best of Humanism. Buffalo: Prometheus, 1988. 224 p. Paper \$14.95.
- GRIMM, ROBERT H., and DANIEL D. MERRILL, eds.: Contents of Thought: Arizona Colloquium in Cognition. Tucson: Arizona UP, 1988. x, 211 p. Cloth \$39.95.
- HANFLING, OSWALD, ed.: Life and Meaning: A Reader. New York: Blackwell, 1987. vii, 255 p. Cloth \$45.00, paper \$16.95.

- HARRISON, ANDREW, ed.: Philosophy and the Visual Arts: Seeing and Abstracting. Boston: Reidel, 1987. xv, 360 p. Cloth \$69.00.
- HATCH, NATHAN O., and HARRY S. STOUT, eds.: Jonathan Edwards and the American Experience. New York: Oxford, 1988. viii, 298 p. Cloth \$29.95.
- HENDLEY, BRIAN P., eds.: Plato, Time, and Education: Essays in Honor of Robert S. Brambaugh. Albany: SUNY, 1988. 326 p. Cloth \$4.50, paper \$14.95.
- HILTON, DENIS J., ed.: Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality. New York: University Press, 1988. xii, 244 p. Cloth \$45.00.
- JAMES, WILLIAM: *Writings* 1902–1910. New York: Library of America, 1988. 1379 p. Cloth \$27.50.
- KAGAN, JEROME, and SHARON LAMB, eds.: The Emergence of Morality in Young Children. Chicago: University Press, 1988. xx, 354 p. Cloth \$24.95.
- KAINZ, HOWARD P., ed.: Philosophical Perspectives on Peace: An Anthology of Classical and Modern Sources. Athens: Ohio UP, 1988. x, 315 p. Cloth \$39.95, paper \$19.95.
- KRELL, DAVID FARRELL, and DAVID WOOD, eds.: Exceedingly Nietzsche: Aspects of Contemporary Nietzsche Interpretation. New York: Routledge & Kegan Paul, 1988. xvi, 179 p. Cloth \$35.00.
- LARSON, GERALD JAMES, and ELIOT DEUTSCH, eds.: Interpreting across Boundaries: New Essays in Comparative Philosophy. Princeton: University Press, 1988. ix, 316 p. Cloth \$38.00.
- McMULLIN, ERNAN, ed.: Construction and Constraint. Notre Dame: University Press, 1988. 250 p. Cloth \$23.95, paper \$11.95.
- McMURRIN, STERLING: The Tanner Lectures on Human Values, vol. VIII. Salt Lake City: Utah UP, 1988. xii, 333 p. Cloth \$25.00.
- MARTIN, LUTHER H., et al., ed.: Technologies of the Self: A Seminar with Michel Foucault. Amherst: Massachusetts UP, 1988. 162 p. Paper \$9.95.
- MATTHEN, MOHAN: Aristotle Today: Essays on Aristotle's Ideal of Science. Alberta, Canada: Academic, 1987. viii, 196 p.
- MENDUS, SUSAN, and DAVID EDWARDS, eds.: On Toleration. New York: Oxford, 1988. 144 p. Cloth \$36.00.
- MILLER, DAVID, et al.: The Blackwell Encyclopedia of Political Thought. New York: Blackwell, 1987. xii, 570 p. Cloth \$60.00.
- MOON, DONALD J., ed.: Responsibility, Rights, and Welfare: The Theory of the Welfare State. Boulder: Westview, 1988. vii, 235 p. Cloth \$38.50, paper \$16.95.
- MULLIGAN, KEVIN, ed.: Speech Act and Sachverhalt: Reinach and the Foundations of Realist Phenomenology. Boston: Nijhoff, 1987. xiii, 344 p. Cloth \$92.50.
- NELSON, JOHN S., et al., eds.: The Rhetoric of the Human Sciences: Language and Argument in Scholarship and Public Affairs. Madison: Wisconsin UP, 1988. xiii, 445 p. Cloth \$24.00.

- PANAGIOTOU, SPIRO, ed.: Justice, Law and Method in Plato and Aristotle. Alberta, Canada: Academic, 1987. 210 p.
- PEACOCKE, ARTHUR, and GRANT GILLET, eds.: Persons and Personality. New York: Blackwell, 1987. viii, 222 p. Cloth \$39.95.
- ROCKMORE, TOM, ed.: Lukacs Today: Essays in Marxist Philosophy. Boston: Kluwer, 1988. 282 p. Cloth \$69.00.
- RUZSA, I., and A. SZABOLCSI, eds.: Proceedings of the '87 Debrecen Symposium on Logic and Language. Budapest: Kiado, 1987. 252 p.
- SCHEFFLER, SAMUEL, ed.: Consequentialism and Its Critics. New York: Oxford, 1988. 294 p. Cloth \$45.00, paper \$14.95.
- SCHOEMAN, FERDINAND, ed.: Responsibility, Character, and the Emotions: New Essays in Moral Psychology. New York: Cambridge, 1988. ix, 358 p. Cloth \$49.50.
- SILVERMAN, HUGH J., and DONN WELTON, eds.: Postmodernism and Continental Philosophy. Albany: SUNY, 1988. 259 p. Cloth \$49.50, paper \$19.95.
- SIMPSON, EVAN: Antifoundationalism and Practical Reasoning: Conversations between Hermeneutics and Analysis. Alberta, Canada: Academic, 1987. vii, 254 p.
- STEIN, GORDON, ed.: A Second Anthology of Atheism and Rationalism. Buffalo: Prometheus, 1988. v, 442 p. Cloth \$22.95.
- TAYLOR, MICHAEL, ed.: Studies in Marxism and Social Theory: Rationality and Revolution. New York: Cambridge, 1988. viii, 271 p. Cloth \$37.50.
- THOMSON, JUDITH JARVIS, ed.: On Being and Saying: Essays for Richard Cartwright. Cambridge: MIT, 1987. viii, 293 p. Cloth \$30.00.
- TYMIENIECKA, ANNA-TERESA, ed.: Analecta Husserliana: The Yearbook of Phenomenological Research, vol. XXII: Morality within the Life- and Social-World. Boston: Reidel, 1987. xv, 586 p. Cloth \$98.00.
- WALDRON, JEREMY, ed.: Nonsense upon Stilts: Bentham, Burke and Marx on the Rights of Man. New York: Methuen, 1987. 236 p. Paper \$13.95.
- WOOD, ALLEN W.: Marx Selections. New York: Macmillan, 1988. viii, 287 p. Paper \$8.00.

NEW BOOKS: TRANSLATIONS AND NEW EDITIONS

- AGACINSKI, SYLVIANE: Aparté: Conceptions and Deaths of Søren Kierkegaard. KEVIN NEWMARK, trans. Tallahassee: Florida UP, 1988. xi, 266 p. Paper \$24.95.
- ARISTOTLE: Poetics with the Tractatus Coislinianus, reconstruction of Poetics II, and the fragments of the On Poets. RICHARD JANKO, trans. Indianapolis: Hackett, 1987. xxvi, 235 p. Cloth \$27.50, paper \$6.95.
- BLANCHOT, MAURICE: Michel Foucault as I Imagine Him. And MICHEL FOUCAULT: Maurice Blanchot: The Thought from Outside. JEFFREY MEHLMAN, trans. Cambridge: MIT, 1988. 109 p. Paper \$16.95.
- BOETHIUS: In Ciceronis Topica. ELEONORE STUMP, trans. Ithaca: Cornell, 1988. xi, 277 p. Cloth \$35.00.

- BORCH-JACOBSEN, MIKKEL: *The Freudian Subject*. CATHERINE PORTER, trans. Stanford: University Press, 1988. ix, 278 p. Cloth \$29.50.
- BUBNER, RÜDIGER: Essays in Hermeneutics and Critical Theory. ERIC MATTHEWS, trans. New York: Columbia, 1988. viii, 262 p. Cloth \$35.00.
- BOSCHETTI, ANNA: The Intellectual Enterprise: Sartre and Les Temps Modernes. RICHARD C. McCLEARY, trans. Evanston, IL: Northwestern, 1988. 279 p. Cloth \$42.95, paper \$14.95.
- COMTE, AUGUSTE: Correspondance Générale et Confessions: Tome VII, 1853-1854. Paulo de Berrêdo Carneiro, ed. Paris: École des hautes études et science sociales, 1987. ci, 326 p.
- DELEUZE, GILLES, and FELIX GUATTARI: A Thousand Plateaus: Capitalism and Schizophrenia. Minneapolis: Minnesota UP, 1987. xix, 610 p. Cloth \$45.00, paper \$17.95.
- DESCARTES: Discours de la Methode, La Dioptrique, Les Meteores, et La Geometrie. France: Fayard, 1986. 476 p.
- DUFRENNE, MIKEL: In the Presence of the Sensuous: Essays in Aesthetics. MARK S. ROBERTS and DENNIS GALLAGHER, trans. Atlantic Highlands, NJ: Humanities, 1987. xxv, 213 p. Cloth \$45.00.
- FICHTE, JOHANN: The Vocation of Man. PETER PREUSS, trans. Indianapolis: Hackett, 1987. xvi, 123 p. Cloth \$18.50, paper \$5.45.
- FUNKE, GERHARD: *Phenomenology: Metaphysics or Method?* Athens: Ohio UP, 1987. xv, 264 p. Cloth \$27.95.
- GADAMER, HANS-GEORG: The Idea of the Good in Platonic-Aristotelian Philosophy. P. CHRISTOPHER SMITH, trans. New Haven: Yale, 1988. xxxi, 182 p. Paper \$8.95.
- GEHLEN, ARNOLD: *Man: His Nature and Place in the World*. CLARE McMILLAN and KARL PILLEMER, trans. New York: Columbia UP, 1987. xxxvi, 420 p. Cloth \$50.00.
- GILSON, ETIENNE: Linguistics and Philosophy: An Essay on the Philosophical Constants of Language. Notre Dame: University Press, 1988. xix, 208 p. Cloth \$29.95.
- HEGEL, G. W. F.: Lectures on the Philosophy of Religion: vol. II: Determinate Religion. PETER C. HODGSON, ed. Berkeley: California UP, 1988. xvi, 825 p. Cloth \$65.00.
- HELVÉTIUS, CLAUDE-ADIREN: De l'Esprit. France: Fayard, 1988. 577 p.
- KIERKEGAARD, SOREN: Either/Or, Parts 1 and 11. HOWARD V. HONG and EDNA H. HONG, eds. and trans. Princeton: University Press, 1987. xix, 701 p. and 528 p. Paper \$13.95 and \$12.95.
- KOSELLECK, REINHART: Critique and Crisis: Enlightenment and the Pathogenesis of Modern Society. Cambridge: MIT, 1988. x, 204 p. Cloth \$22.50.
- LE ROY, LOYS: De la Vicissitude ou Variété des Choses en l'Univers. France: Fayard. 1988. 443 p.
- MABLY, GABRIEL BONNOT: De l'Étude, de l'Histoire, suivi de la Manière d'Ecrire l'Histoire. France: Fayard; 1988. 405 p.

RICOEUR, PAUL: Time and Narrative: vol. III. KATHLEEN BLAMEY and DAVID PELLAUER, trans. Chicago: University Press, 1988. 355 p. Cloth \$29.95.

SCHLICK, MORITZ: The Problems of Philosophy in Their Interconnection: Winter Semester Lectures 1933–34. HENK L. MULDER, A. J. KOX, and RAINER HEGSELMANN, eds. Boston: Reidel, 1987. x, 213 p. \$89.00.

sloterdijk, Peter: Critique of Cynical Reason. Minneapolis: Minnesota UP, 1988. xxxix, 558 p. Cloth \$45.00, paper \$19.95.

VEYNE, PAUL: Did the Greeks Believe in Their Myths? An Essay on the Constitutive Imagination. PAULA WISSIG, trans. Chicago: University Press, 1988. xii, 161 p. Cloth \$25.00, paper \$10.95.

ON PROOF FOR THE EXISTENCE OF GOD, AND OTHER REFLECTIVE INQUIRIES

PAUL VJECSNER

A Cartesian reexamination of basic presuppositions, old and new, in philosophy and sciences. The treatise challenges many assertions of fact in these, as well as rising convictions that many truths are indeterminable. Answers are offered thus to questions of meaning, free will, forces in physics, axioms and paradoxes in logic and mathematics, transcendent realities, and so forth. Correspondingly the author does also not depend on methods of inquiry in use, reverting to exposition in commonly comprehended form.

ON PROOF FOR THE EXISTENCE OF GOD, AND OTHER REFLECTIVE INQUIRIES

BY PAUL VIECSNER

PENDEN

264 pages, 85 diagrams, index ISBN 0-9619519-0-7 Cloth bound, \$20.00 Send order for "Reflective Inquiries" to PENDEN, P.O. Box 464, New York, N.Y. 10101

Theory and Decision

An International Journal for Methods and Models in the Social and Decision Sciences

Editor-in-Chief

Prof. Bertrand Munier, Université d'Aix-Marseille III, France

The field of decision in the social sciences has been investigated from many aspects. All too often, however, different research approaches have been followed separately by management science, economics, psychology, operations research, praxeology, cybernetics and computer aided decision-making.

They need mutual enrichment. *Theory and Decision* addresses the cross-fertilization of the methods and models of these disciplines. The purpose of the journal is to let the engineering of decision-making (i.e. of intelligence, choice, sorting out, ranking, uncertainty and conflict resolution), gradually emerge.

Subscription Information

ISSN 0040-5833

1988. Volumes 24-25 (6 issues)

Institutional rate: Dfl. 412.00/US\$194.00 incl. postage/handling Private rate: Dfl. 160.00/US\$69.00 incl. postage/handling

Private subscriptions should be sent direct to the publishers D. Reidel Publishing Company



Kluwer Academic Publishers



P.O. Box 989, 3300 AZ Dordrecht, The Netherlands 101 Philip Drive, Norwell, MA 02061, U.S.A.

Erkenntnis

An International Journal of Analytic Philosophy

Editors

Carl G. Hempel, Princeton University, USA; Wolfgang Stegmüller, Universität München, FRG, Wilhelm K. Essler, Universität Frankfurt, FRG

Erkenntnis is a philosophical journal publishing papers on foundational studies and scientific methodology covering:

'Philosophy of Science' and 'Analytic Philosophy'; philosophy of language, of logic, and of mathematics; foundational problems of physics and of other natural sciences; foundations of normative disciplines such as ethics, philosophy of law and of aesthetics; methodology of the social sciences and the humanities; history of scientific method. *Erkenntnis* was first published in 1930, and succeeded *Annalen der Philosophie*, which appeared from 1919–29.

Subscription Information

1988, Volume 28-29 (6 issues)

ISSN 0165-0106

Institutional rate: Dfl. 452.00/US\$214.00 incl. postage/handling Private rate: Dfl. 172.00/US\$72.00 incl. postage/handling



Kluwer Academic Publishers

P.O. Box 989, 3300 AZ Dordrecht, The Netherlands 101 Philip Drive, Norwell, MA 02061, U.S.A.



CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 9, SEPTEMBER 1988

JUSTICE, CARE, GENDER BIAS

AROL GILLIGAN poses two separable, though in her work not separate, challenges to moral theory. The first is a challenge to the adequacy of current moral theory that is dominated by the ethics of justice.1 The ethics of justice, on her view, excludes some dimensions of moral experience, such as contextual decision making, special obligations, the moral motives of compassion and sympathy, and the relevance of considering one's own integrity in making moral decisions. The second is a challenge to moral theory's presumed gender neutrality. The ethics of justice is not gender neutral, she argues, because it advocates ideals of agency, moral motivation, and correct moral reasoning which women are less likely than men to achieve; and because the moral dimensions excluded from the ethics of justice are just the ones figuring more prominently in women's than men's moral experience.

The adequacy and gender bias charges are, for Gilligan, linked. She claims that the ethics of justice and the ethics of care are two different moral orientations. Whereas individuals may use both orientations, the shift from one to the other requires a Gestalt shift, since "the terms of one perspective do not contain the terms of the other" (ibid., p. 30). The exclusion of the care perspective from the ethics of justice simultaneously undermines the adequacy of the

² Carol Gilligan, "Moral Orientation and Moral Development," in Women and Moral Theory, Eva Feder Kittay and Diana T. Meyers, eds. (Totowa, NJ: Rowman &

In referring to the 'ethics of justice' and the 'ethics of care', I do not assume that either one is some monolithic, unified theory; rather, I use these terms, as Gilligan suggests, to designate different orientations—loosely defined sets of concepts, themes, and theoretical priorities—which we understand sufficiently well to pick out who out who is speaking from which orientation, but which are not so rigid as to preclude a great deal of disagreement within each orientation.

ethics of justice (it cannot give a complete account of moral life) and

renders it gender-biased.

Some critics have responded by arguing that there is no logical incompatibility between the two moral orientations.³ Because the ethics of justice does not in principle exclude the ethics of care (even if theorists within the justice tradition have had little to say about care issues), it is neither inadequate nor gender-biased. Correctly applying moral rules and principles, for instance, requires, rather than excludes, knowledge of contextual details. Both orientations are crucial to correct moral reasoning and an adequate understanding of moral life. Thus, the ethics of justice and the ethics of care are not in fact rivaling, alternative moral theories. The so-called ethics of care merely makes focal issues that are already implicitly contained in the ethics of justice.

Suppose the two are logically compatible. Would the charge of gender bias evaporate? Yes, so long as gender neutrality only requires that the ethics of justice could, consistently, make room for the central moral concerns of the ethics of care. But perhaps gender neutrality requires more than this. Since the spectre of gender bias in theoretical knowledge is itself a moral issue, we would be well advised to consider the question of gender bias more carefully before concluding that our moral theory speaks in an androgynous voice. Although we can and should test the ethics of justice by asking whether it could consistently include the central moral issues in the ethics of care, we might also ask what ideologies of the moral life are likely to result from the repeated inclusion or exclusion of particular topics in moral theorizing.

Theorizing that crystallizes into a tradition has nonlogical as well as logical implications. In order to explain why a tradition has the contours it does, one may need to suppose general acceptance of particular beliefs that are not logically entailed by any particular theory and might be denied by individual theorists were those beliefs articulated. When behavioral researchers, for example, focus almost exclusively on aggression and its role in human life, neglecting other behavioral motives, their doing so has the nonlogical implication that aggression is, indeed, the most important behavioral motive. This is

³ The logical compatibility thesis is, for example, advanced by Jean Grimshaw, *Philosophy and Feminist Thinking* (Minneapolis: Minnesota UP, 1986); Owen Flanagan and Kathryn Jackson, "Justice, Care, and Gender: The Kohlberg-Gilligan Debate Revisited," *Ethics*, XCVII (1987): 622–637; Thomas E. Hill, Jr., "The Importance of Autonomy," in *Women and Moral Theory*; George Sher, "Other Voice, Other Rooms? Women's Psychology and Moral Theory," in *Women and Moral Theory*.

Digitized by Arya Samaj Foundation Chennai and eGangotri because only a belief like this would explain the rationality of this pattern of research. Such nonlogical implications become ideologies when politically loaded (as the importance of aggression is when coupled with observations about women's lower level of aggression).

When understood as directed at moral theory's nonlogical implications, the gender-bias charge takes a different form. Even if the ethics of justice could consistently accommodate the ethics of care. the critical point is that theorists in the justice tradition have not said much, except in passing, about the ethics of care, and are unlikely to say much in the future without a radical shift in theoretical priorities; and concentrating almost exclusively on rights of noninterference, impartiality, rationality, autonomy, and principles creates an ideology of the moral domain which has undesirable political implications for women. This formulation shifts the justice-care debate from one about logical compatibility to a debate about which theoretical priorities would improve the lot of women.

I see no way around this politicization of philosophical critique. If we hope to shape culture, and not merely to add bricks to a philosophical tower, we will need to be mindful of the cultural/political use to which our thoughts may be put after leaving our wordprocessors. This mindfulness should include asking whether our theoretical work enacts or discredits a moral commitment to improving the lot of women.

Starting from the observation that the ethics of justice has had centuries of workout, I want to ask what ideological implications a concentration on only some moral issues might have and which shifts in priorities might safeguard against those ideologies. This particular tack in trying to bring the ethics of care to center stage has the double advantage of, first, avoiding the necessity of making charges of conceptual inadequacy stick, since it does not matter what the ethics of justice could consistently talk about, only what it does talk about; and, second, of avoiding the question of what, from an absolute, ahistorical point of view moral theory ought to be most preoccupied with.

The following reflections on moral theorizing about the self, knowledge, motivation, and obligations are not meant to be exhaustive but only to suggest some reasons for taking the charge of gender bias in ethics seriously. I shall sometimes stray rather far afield from the ethics of care, since my aim is not to defend the ethics of care but to advocate some shifts in theoretical priorities.

I. THE MORAL SELF

One concern of moral theory has been with broadening our sensitivities about who has morally considerable rights and interests. The

ordinary individual confronts at least two obstacles to taking others' rights and interests seriously. One is his own self-interest, which inclines him to weigh his own rights and interests more heavily; the other is his identification with particular social groups, which inclines him to weigh the rights and interests of co-members more heavily than those of outsiders. Immanuel Kant had a lot to say about the former obstacle, David Hume about the latter. Sensitivity to our failure to weigh the rights and interests of all members of the moral community equally led moral theorists to focus, in defining the moral self, on constructing various pictures of the moral self's similarity to other moral selves in an effort to underscore our common humanity and thus our entitlement to equal moral consideration. Kant's identification of the moral self with the noumenal self, thus minus all empirical individuating characteristics, is one such picture. Emphases on shared human interests in life, health, etc., serve a similar purpose. And so does John Rawls's invocation of a "veil of ignorance."

Providing us with some way of envisioning our shared humanity, and thus our equal membership in the moral community, is certainly an important thing for moral theory to do. But too much talk about our similarities as moral selves, and too little talk about our differences has its moral dangers. For one, unless we are also quite knowledgeable about the substantial differences between persons, particularly central differences due to gender, race, and class, we may be tempted to slide into supposing that our common humanity includes more substantive similarities than it does in fact. For instance, moral theorists have assumed that moral selves have a prominent interest in property and thus in property rights. But property rights may have loomed large on the moral horizons of past moral theorists partly, or largely, because they were themselves propertied and their activities took place primarily in the public, economic sphere. Historically, women could not share the same interest in property and concern about protecting it, since they were neither legally entitled to hold it nor primary participants in the public, economic world.4 And arguably, women do not now place the same priority on property. (I have in mind the fact that equal opportunity has had surprisingly little impact on either sex segregation in the workforce or on women's, but not men's, accommodating their work and work schedules to

⁴ Annette Baier makes a similar point in "Trust and Antitrust," *Ethics*, XCVI (1986): 231–260. There she argues that understanding moral relations in terms of contracts and voluntary promises reflects the social lives of male moral theorists: "Contract is a device for traders, entrepreneurs, and capitalists, not for children, servants, indentured wives, and slaves" (p. 247).

s'

h

es

ly

ıe

ır

al

al

0

ty

ll

1-

ar

y,

ıt

r-

l-

1-

e

al

n

re

r

n

it 1-

re

le

s,

0

VI

ts:

Digitized by Arya Samaj Foundation Chennai and eGangotri

childrearing needs. One explanation is that income matters less to women than other sorts of considerations. The measure of a woman, unlike the measure of a man, is not the size of her paycheck.) Seyla Benhabib⁵ summarizes this point by suggesting that a singleminded emphasis on common humanity encourages a "substitutionalist universalism" where universal humanity "is defined surreptitiously by identifying the experiences of a specific group of subjects as the paradigmatic case of all humans" (*ibid.*, p. 158).

In addition to encouraging us to overlook how our basic interests may differ depending on our social location, the emphasis on common humanity, because it is insensitive to connections between interests, social location, and power, deters questions about the possible malformation of our interests as a result of their development within an inegalitarian social structure. Both dangers plague the role-reversal test, some version of which has been a staple of moral theorizing. Although the point of that test is to eliminate egoistic bias in moral judgments, without a sensitivity to how our (uncommon) humanity is shaped by our social structure, role-reversal tests may simply preserve, rather than eliminate, inequities. This is because rolereversal tests either take individuals' desires as givens, thus ignoring the possibility that socially subordinate individuals have been socialized to want the very things that keep them socially subordinate (e.g., Susan Brownmiller⁶ argues that women have been socialized to want masochistic sexual relations); or, if they take into account what individuals ought to want, role-reversal tests typically ignore the way that social power structures may have produced an alignment between the concept of a normal, reasonable desire and the desires of the dominant group (so, for example, much of the affirmative action literature takes it for granted that women ought to want traditionally defined male jobs with no consideration of the possibility that women might prefer retailoring those jobs so that they are less competitive, less hierarchical, and more compatible with family responsibilities).

In short, without adequate knowledge of how very different human interests, temperaments, lifestyles, and commitments may be, as well as a knowledge of how those interests may be malformed as a result of power inequities, the very egoism and group bias that the focus on common humanity was designed to eliminate may slip in as a result of that focus.

b "The Generalized and Concrete Other: The Kohlberg-Gilligan Controversy and Moral Theory," in Women and Moral Theory, p. 158.

Against Our Will: Men, Women, and Rape (New York: Simon & Schuster, 1975)

The objection here is not that a formal, abstract notion of the moral self's common humanity is wrong and ought to be jettisoned. Nor is the objection that a formal notion of the moral self logically entails a substitutionalist universalism. The objection is that repetitive stress on shared humanity creates an ideology of the moral self: the belief that our basic moral interests are not significantly, dissimilarly, and sometimes detrimentally shaped by our social location. Unless moral theory shifts its priority to knowledgeable discussions of human differences—particularly differences tied to gender, race, class, and power—lists and rank orderings of basic human interests and rights as well as the political deployment of those lists are likely to be sexist, racist, and classist.

II. MORAL KNOWLEDGE

Central to moral theory has been the issue of how moral principles. and hence moral decisions in particular cases, are to be justified. We owe that interest in justification in large measure to the modern period's concern to find foundations for knowledge that are, in principle, accessible to any rational individual. The concern with justifying moral knowledge meant that some questions, but not others, were particularly important for moral theory to address. First, how should an adult who has acquired a wide variety of moral views as a result of his socialization into a particular cultural tradition go about evaluating those views? That is, how do we distinguish mere inherited prejudices from legitimate moral beliefs? Second, given that the correctness of particular moral judgments depends, in part, on the correctness of the general moral principles that we bring to bear on particular cases, how do we justify those general moral principles? In either case, the answer involves showing that our moral views and principles can survive various tests of rationality, e.g., that they are consistent and universalizable.

The danger of asking these questions lies not in their being the wrong questions, but in their being only some of the right questions for a moral epistemology. As adults, moral theorists may naturally find questions about distinguishing learned prejudices from justified moral beliefs more pertinent to their own lives. And certainly one of the capacities that we hope moral agents will acquire is the capacity to draw just those kinds of distinctions. But we may pay a price by too strongly emphasizing the acquisition of moral knowledge through individual, adult reflection. For one, this emphasis contributes to the idea that the self who is capable of moral knowledge is, in Benhabib's caustic words, "a mushroom behind a veil of ignorance"; that is, that the moral knower, like a mushroom, has neither mother nor father, nor childhood education (op. cit, p. 166). Thus, we may lose sight of

the fact that our adult capacity for rational reflection, the size of our adult reflective task, and quite possibly our motivation to act on reflective judgments depend heavily on our earlier moral education. Whereas moral theory has not been altogether blind to the importance of moral education, few have given moral education a role comparable to that of adult reflection in the acquisition of moral knowledge. (Francis Hutcheson comes to mind as a notable exception.) The result is an ideology of moral knowledge: the belief that moral knowledge is not only justified but also acquired exclusively or most importantly through rational reflection. Women have special reason to be concerned about this ideology. Women's traditional role has included the moral education of children. The significance of women's work in transmitting moral knowledge and instilling a moral motivational structure (either well or poorly) is likely to remain invisible so long as the theoretical focus remains on adult acquisition of moral knowledge.

1-

1.

ıs

ts

n

n

h

ot

n

e

n

t,

0

ır

ie

d

эf

ty

0

ŗh

1e

's

at

r,

of

More importantly, stressing the corrective efficacy of individual, rational reflection creates a second ideology of moral knowledge: the belief that individual reflection, if it conforms to cannons of rationality, guarantees the truth of one's moral judgments. It is untrue that any rational individual who applies sufficient reflective elbow grease can adequately assess the justifiability of his moral views or go behind a veil of ignorance and come out with the correct moral principles. Our being motivated to raise questions of justification in the first place and our ability to address those questions once raised depends at least partially on the social availability of moral criticisms and of morally relevant information. The nineteenth century's moral injunction against women's pursuing advanced education was not simply the product of failed rational reflection. It was tied, on the one hand, to a societal assumption that women's unequal status was morally unproblematic; and, on the other hand, to medical misinformation about the connection between women's intellectual activity and the healthy functioning of their reproductive organs. Thus, moral questions about the policy of barring women from higher education were unlikely to be raised, since rationality does not require indiscriminately questioning any and all policies but only those reasonably open to question. Even if raised, they were unlikely to be answered in women's favor, since, at the time, there appeared to be morally relevant differences between men and women. Only women could harm themselves and produce mentally and physically defective children dren as a result of education.

Without an equal theoretical stress on the social determinants of moral knowledge—particularly the potential alignment of moral and

factual beliefs with social power structures—the very reflective processes that were designed to criticize cultural prejudices may simply repeat those prejudices. In emphasizing moral interdependency over moral autonomy, the ethics of care provides the kind of theoretical focus that could make moral education and the social determinants of moral knowledge salient.

III. MORAL MOTIVATION

Moral theorizing, particularly though not exclusively in the Kantian tradition, has focused on the motivating role of thoughts of duty, of what is right or what contributes to general happiness. Moral action should stem from a regard for morality itself rather than from non-moral thoughts, self-interest, or happily altruistic emotions, since only a regard for morality itself provides a reliable spur to moral action, and only a regard for morality focuses our attention on the kinds of considerations that ensure right action.

One bone of contention in the justice-care debate has been over whether the requirement to have duty as one's motive necessarily excludes being motivated by care, sympathy, compassion, or the personally involved motives of love, loyalty, and friendship. Marcia Baron⁷ has argued forcefully for the compatibility of duty with more personally involved motives. Central to her argument is the distinction between primary and secondary motives.

A primary motive supplies the agent with the motivation to do the act in question, whereas a secondary motive provides limiting conditions on what may be done from other motives. Although qua secondary motive it cannot by itself move one to act, a secondary motive is nonetheless a motive, for the agent would not proceed to perform the action without the "approval" of the secondary motive (*ibid.*, p. 207).

Being motivated by duty as a secondary motive amounts to no more than the realization that one would not act on one's love or one's compassion if doing so conflicted with what morally ought to be done. Thus, being morally motivated by duty does not require taking an emotionally uninvolved, alienated stance toward others. It merely requires cautious willingness to refrain from action that conflicts with what one ought to do. Moreover, Baron argues that doing what one ought to do may well include cultivating one's capacity for sympathy and compassion, since merely "going through the motions" is often less than what duty requires.

I find Baron's argument convincing, and truer to Kant. But, even if a duty-centered ethics can consistently accommodate caring atti-

^{7 &}quot;The Alleged Moral Repugnance of Acting from Duty," this JOURNAL, LXXXI, 4 (April 1984): 197–220.

tudes, one can still object that the repeated opposition of duty to self-interest creates an ideology of moral motivation: the belief that we are psychologically so constructed that duty must usually supply a primary motive.

Almost invariably in moral theory, it is the lack of other-directed attitudes which is cited as the largest motivational obstacle to doing what morality requires. Agents find it difficult to behave morally because (1) they are egoistic and are inclined initially to be motivated by self-interest and to weigh their own interests more heavily than others'; and (2) even where they stand to gain nothing by acting immorally, they are initially indifferent to others' welfare. Moral thoughts, particularly the thought of duty, combat egoism and indifference by supplying a primary motive to do what morality requires which we otherwise would lack. Thus, moral theory constructs an image of the moral agent as psychologically so constituted as to require that duty be his primary motive. Conceding, in the way Baron does, that duty may operate as a secondary motive in some people or on some occasions does nothing to counter this image of the agent's psychology.

The narratives of the women in Gilligan's study, however, suggest a very different motivational picture for women.8 At the earliest stage of moral development, women may share this egoistic psychology. But, at later stages, it is an unreflective and often self-excluding sympathy for others which poses the main motivational obstacle. Far from the lure of self-interest, the motivational problem for adult women is how to place proper limits on the inclination to respond to others' needs. The problem is not one of getting duty to operate as a primary motive, but of how to get it to function properly as a secondary motive. Moral theories that emphasize conquering self-interest by cultivating a sense of duty (or by cultivating sympathy) only reinforce women's inclination to act on caring attitudes unchecked by considerations of duties to self or overriding duties to others. If women's elective underparticipation in the workforce, overassumption of familial duties, and nonreportage of date rape and marital abuse concern us morally and politically, we might do well to shift theoretical priority from the conflict between duty and self-interest to that between duty and care.

IV. MORAL OBLIGATIONS

The concern of traditional moral theory with impartiality emerges variously out of a concern with countering self-interest, enlarging our sentiments, and introducing greater consistency into moral

 $^{^{8}}$ In a Different Voice (Cambridge: Harvard, 1982).

judgments. This, too, has been a bone of contention in the justice-care debate: Does a fully impartial ethics necessarily exclude special moral obligations to friends and family and thus exclude considerations that are more likely to figure centrally in women's actual moral thinking given their traditional and ongoing familial role? In defending the adequacy of the ethics of justice, George Sher⁹ argues from a Rawlsian contractarian point of view that the selection of impartial rules may well include selecting rules that dictate special obligations.

The contractors' ignorance does rule out the choice of principles that name either specific agents who are allowed or required to be partial or the specific recipients of their partiality. However, this is irrelevant; for the question is not whether any given person may or should display partiality to any other, but rather whether all persons may or should be partial to their wives, husbands, or friends. The relevant principles, even if licensing or dictating partiality, must do so impartially. Hence, there is no obvious reason why such principles could not be chosen even by contractors ignorant of the particulars of their lives.

Similarly, Kantians might argue from the idea of implicit promises made to family or friends, and utilitarians might argue from considerations of maximizing welfare, that it is possible to give preference to friends and families without giving up the idea that no one counts for more than one. I want to concede this point that the ethics of justice leaves logical room for special obligations.

But, when moral theory is largely silent about special obligations or brings them in as addenda, two ideologies of moral life get created: the first is the belief that it is self-evident that general obligations are morally more important than special obligations. This ideology is troubling, because the division of the moral world into general obligations governing public relations with relative strangers and special obligations governing personal relations with family and friends so closely parallels the genderized division of spheres into public and private. The value of women's private domestic work has been too quickly dismissed in the past by those who assumed that public productive labor is self-evidently more important than private reproductive labor. One might, then, reasonably worry about the way moral theory, perhaps inadvertently, confirms this quick dismissal of the private realm as "of course" less important.

The second ideology created by the repeated focus on general obligations is the belief that general obligations are experientially more frequently encountered; they deserve more attention because

⁹ "Other Voices, Other Rooms?" op. cit., p. 186.

questions about them come up more often. Women, however, in addition to typically being more involved in familial caretaking, overwhelmingly dominate service and caring jobs; and the interpersonal relationships in those jobs bear many of the same characteristics as do private, familial relationships. They are often ongoing, dependency relations and/or involve heightened expectations that the worker will have a special concern for and advocacy relation to the client/employer (e.g., teaching, daycare, nursing, social work, secretarial work, and airline stewardessing). Even in traditionally male iobs, both employers and clients may expect, in virtue of women's caring social role, more from women workers (for example, to be warmer and more supportive) than general moral obligations require. Given these kinds of considerations, theoretical emphasis on general obligations (which would incline one to think that special obligations are experientially less frequently encountered) quite naturally evokes the question "Whose moral experience is being described?" Moreover, so long as moral theory continues to depict public moral relations as though they were governed almost exclusively by general obligations, which leave a good deal of latitude for the pursuit of self-interest, we are unlikely to see that women's public moral lives, not just their private ones, leave less scope for the pursuit of self-interest than men's.

THE CHARGE OF GENDER BIAS

I have argued that repeated focusing in moral theorizing on a restricted range of moral problems or concepts produces ideologies of the moral life which may infect our philosophical as well as our popular, cultural beliefs. I want to emphasize that this results from the cumulative effect of moral theorizing rather than from errors or omissions in particular ethical works considered individually. I also want to re-emphasize that those ideologies need have been neither explicitly articulated nor believed by any serious moral philosopher (though some surely have). They are, rather, "explanatory beliefs" whose general acceptance would have to be supposed in order to explain the rationality of the particular patterns of philosophical conversation and silence which characterize moral theory. The charge of gender bias is thus not addressed to individual thinkers so much as to the community of moral theorists or, alternatively, to a tradition of moral theorizing.

The call for a shift in theoretical priorities is simultaneously a call for a shift in our methods of evaluating moral theories. Evaluation is not exhausted by carefully scrutinizing individual theories, since in the process of theorizing in a philosophical community we unavoidably contribute to the establishment of a tradition of moral thinking

which may implicitly, in virtue of common patterns of talk and silence, endorse views of the moral life which go beyond those of individual contributors. The nonlogical implications of theorizing

patterns require evaluation as well.

But, if moral theory suffers from a lopsidedness that produces ideologies of the moral life, why be particularly concerned with eliminating gender bias? Would not the more basic, and broader, philosophical task be to eliminate bias in general? Would not a bias sensitive (but gender insensitive) critique do all the work? There is indeed an interesting coincidence between the critiques stemming out of Gilligan's work with critiques having no clear connection to it or feminist theory. 10 The call, coming out of the ethics of care, for a de-emphasis on the role of reflective, principled reasoning curiously coincides with an independent resuscitation of virtue ethics. Similarly, Gilligan's attention to personal integrity coincides with comparable but independent worries about the threats posed by an overly demanding moral theory to personal integrity. The same is true of philosophical demands for moral attention to the good life, compassion, and special obligations. Thus, sensitivity to gender issues would seem unnecessary for philosophical critiques whose consequence, though not intent, would be a gender neutral moral theory.

Perhaps, but I suspect not. Some moral issues are arguably more critical for women, and thus achieving gender neutrality is partly a matter of prioritizing those issues. But eliminating gender bias cannot be equated (though possibly reducing gender bias can) with simply prioritizing those "women's issues" irrespective of the content of the analysis of those issues. These same issues also have a place in men's moral experience. For that reason, male moral philosophers too may have cause to regret moral theory's neglect of special relations, virtue ethics, compassion and the problem of limiting compassionate impulses; and it is thus no surprise that some of the same critiques of moral theories are coming from both feminist and nonfeminist quarters. But, given that our lives are thoroughly genderized, there is no reason to suppose that gender bias cannot recur in the discussion of these "women's moral issues." Which virtues, after all, will we make focal—intellectual virtues or inter-

¹⁰ I have in mind Bernard Williams, "Persons, Character, and Morality" and "Moral Luck," in *Moral Luck* (New York: Cambridge, 1981); and "Integrity," in *Utilitarianism: For and Against*, J. J. C. Smart and Bernard Williams, eds. (New York: Cambridge, 1973); Susan Wolf, "Moral Saints," this JOURNAL, LXXIX, 8 (August 1982): 419–429; Andrew Oldenquist, "Loyalties," this JOURNAL, LXXIX, 4 (April 1982): 173–193; Lawrence Blum, *Friendship*, *Altruism*, and *Morality* (London: Routledge & Kegan Paul, 1980).

personal virtues? And what will we say about individual virtues? Will we, as Annette Baier does (op. cit.), examine how virtues may undergo deformation in different ways depending on our place in power structures? Which kind of compassion will become paradigmatic: the impersonal, public compassion for strangers and unfortunate populations, or the personal, private compassion felt for friends, children, and neighbors? Will we repeat the same militaristic metaphors of conquest and mastery in describing conflicts between compassion and duty which have dominated descriptions of the moral agent's relation to his self-interest? And, in weighing the value of personal integrity against the moral claims of others, will we take into account the way that gender roles may affect both the value we attach to personal integrity and the weight we attach to others' claims?

The possibility of gender bias recurring in the process of redressing bias in moral theory derives from the fact that philosophical reasoning is shaped by extra-philosophic factors, including the social location of the philosophic reasoner and his audience as well as the contours of the larger social world in which philosophic thought takes place. It is naive to suppose that a reflective, rational, but gender-insensitive critique of moral theory will have the happy outcome of eliminating gender bias. So long as we avoid incorporating gender categories among the tools for philosophical analysis, we will continue running the risks, whether we work within or counter to the tradition, of importing gender bias into our philosophical reflection and of creating an ideology of the moral life.

CHESHIRE CALHOUN

College of Charleston

AN EXPLANATION OF RETRIBUTION

■ HE universal insistence upon retribution for grievous crimes is deeply felt, intractable, and largely independent of utilitarian considerations. The pursuit of Adolph Eichmann, Josef Mengele, and other Nazis in their dotage, tending their rose gardens in South America, makes no utilitarian sense whatever. They will not do their crimes again, nor would their punishment deter others. On this issue the great majority of us find our God in the Old Testament, not the New. No one who reads accounts of courtroom reactions of crime victims, or of the parents of children whose murderers have been convicted (or acquitted), can doubt that the world demands retribution for criminal harm—not just compensation or restitution, which often is impossible, but retribution. After every conviction the victims, or their relatives, applaud, or cry with relief, and otherwise indicate that the world could never be right again until the one who hurt them so terribly received his due. Some of these people will call it justice, some will defiantly admit it is revenge. But, if these courtroom responses lack moral significance and are mere effusions of vengeance, so too does the 40-year search for Josef Mengele and other Nazis. Whether retribution nonetheless has utility is an issue I shall pursue further on.

Whereas the desire for retribution does not prove it right, I believe there are arguments that explain the function and nature of retribution in such a way that together they provide some considerations in favor of morally accepting retributive punishment. The first argument aims to establish the social need for retribution, on the grounds that humans necessarily live in moral communities and the existence of moral communities depends on the acceptance of retributive justice. The second argument rebuts the objection that an insistence on retribution is a mere desire for revenge and not a moral claim. This argument aims to show that retributive justice is sanitized revenge, that is, that revenge becomes retributive justice when certain empirical social conditions hold. After these arguments are laid out, we need to take up the possibility that retributive judgments might still be false moral judgments, even if it is established that they are moral judgments and are socially necessary; and a final task is to examine utilitarian grounds for retaining the social institution of retribution. I. MORAL COMMUNITIES

The first of these two arguments is the following:

(1) Humans are innately social animals who can flourish and achieve their full humanity only in society.

0022-362X/88/8509/0464\$01.50

© 1988 The Journal of Philosophy, Inc.

(2) A human society is a moral community.

- (3) A moral community is such that members hold one another personally accountable for harm to fellow members and to the common good.
- (4) To hold persons personally accountable for harm to fellow members and to the common good is to consider them deserving of blame and punishment.
- (5) To consider persons deserving of blame and punishment is morally to accept retribution.

Therefore:

(6) Humans can flourish and achieve their full humanity only if they morally accept retribution.

Step (1) is empirical, except for debate about achieving full humanity. This is intended to be a very strong claim; a way to assert it without the quasi-evaluative language is that breeding populations of humans survive only in societies. But, even if they could physically survive outside society, their lives would be so stunted that we would not recognize them as human. Most anthropologists and students of human evolution maintain that our ancestors were social animals throughout the Pleistocene, before we were even human. This is to say we are innately social, not social by convention or a "social contract," and that we evolved the emotional equipment for cooperative social living.

Extended infancy, pair bonding, and loss of estrus are evolutionary developments of hominids as they became social animals; and the evolution of these traits probably was reciprocally reinforced by early social life. Collaborative hunting and gathering, tool use, language, the need for familiar people, territory, and routines, the universal fear of being outcast, and the world-wide development of ritual and ceremony are among the things constitutive of human sociality. I hypothesize also that humans became genetically primed to need the socialization process itself, and not just its benefits. As Mary Midgley put it, humans come half finished, requiring a culture to complete them.

By a moral community (in step 2), I mean a society that includes the following features.

(a) Members have a sense of societal identity and social boundaries, and are able to distinguish members from nonmembers. Unless alienated, their sense of who they are is partly constituted by their social identity.

¹ Beast and Man (Ithaca: Cornell, 1978), p. 286.

(b) There are personal virtues, traditions and ceremonies, and rules of social morality that are shared and consciously transmitted from generation to generation.

(c) The community is, to its members, a noninstrumental value, a common good that is more than a mere protective association or

interest group.

(d) Members have group loyalty as well as impersonal morality, and are inclined to be proud when their community prospers, ashamed when it is disgraced, and indignant when it is harmed.

In order to live together cooperatively people must be socialized. We want the behavior of people we meet to be reasonably predictable and nonpredatory. A human community, unlike (but not completely unlike) an ant colony or a herd of gazelle, has social morality and group loyalty because we can conceptualize, name, and teach the dispositions for social cooperation. We also can figure out the advantages of being free riders, which requires social countermeasures. A moral community is just the form a herd or colony takes, given a species with the innate sociality, intelligence, and language ability that characterize humans.

A moral community is distinct from an interest group or protective association. People may unite to achieve a common end such as defeating a tax bill. They form an interest group; its value is wholly instrumental and, if it fails, its reason for existence is gone. An interest group makes no claim on our loyalty and one cannot be alienated from it. But an interest group might evolve into a moral community; then it is an "organic" community, its good being more than instrumental to its members. In this case one of one's social identities derives from it, and one can be proud or ashamed, loyal or disloyal, or alienated from it.

In step (3), which says that in a moral community members hold one another personally accountable for their behavior, one could define 'personally accountable' as causally accountable and such that incarceration would be utile. On this interpretation, step (3) would not imply step (4)—that some people are considered to deserve blame and punishment; or, if it did imply (4), it would be in a sense of 'blame' and 'punishment' such that (4) did not imply (5)—that one morally accepts retribution. I mean 'personally accountable' in a sense that makes (4) and (5) true by definition. Thus, (3) is the crucial premise, so let us now turn to it.

A sense of belonging to a community depends on an ability to individuate it from others, which in turn depends on being able to tell members from nonmembers. One way is for members to wear blue feathers. A more important way, in part definitive of the differ-

ence between members and nonmembers, is in terms of who is held personally accountable for harm and expected to make sacrifices for the common good. This assignment of responsibility is complex, depending on the degree of harm done and on social distance. For example, outsiders are not held accountable for careless criticism or for the neglect of my club, neighborhood, church, or labor union; but almost any normal human is held accountable for arson or robbery. I say 'almost,' because those whose social distance is great, such as Amazonian tribe members, often are not held accountable. Retributive thoughts and feelings do not arise unless the criminal is in some sense one of our own, someone from whom we expect compliance and group regard; otherwise our thoughts remain utilitarian and do not go beyond self-defense or quarantine. We are indignant and angry when the transgressor belongs to our society (in some cases just being a member of a nonprimitive human society is enough), and retribution is the expression of this indignation.

A society that never held its members personally accountable for good or evil, never expressed its collective will through praise and censure, would have a weak conception of the difference between members and nonmembers; and, what may amount to the same thing, it would have a weak or diffuse conception of a common good worth defending. It is doubtful we could possess common values and a common way of life if we were totally lacking in indignation at affronts to them. Serious crimes, when they go unpunished, diminish the value we place on our social identities, and hence our valuation of ourselves

Only if I regard a community as mine am I capable of pride, shame, or indignation regarding it. In general, membership in a moral community is revealed by who is affronted by insults to it, who enlists, or puts in time and contributes money, by who is blamed and criticized for not behaving in these ways, or feels shame and guilt. It follows that a powerful way to tell young delinquents they still belong and are members of a moral community is to hold them personally accountable and punish them, whereas not holding them accountable is to alienate them more than they already are, to cast them out emotionally. Criminal punishment that is not done on openly retributive grounds exacerbates a criminal's alienation from mainstream society.

A moral community is an intricate network of felt obligations and expectations. Each of us has a number of social roles and the viability of society requires reliance on others for the behavior expected of a parent, teacher, student, passerby, neighbor, driver, business person, worker of any kind, and so on. We constantly judge the perfor-

mance of the people we encounter, and when it falls short, as everyone's does occasionally, we criticize them, hold them responsible.

What I am suggesting is that retributive criminal justice and the sense of mutual responsibility found in well-functioning moral communities are two modes of a common moral sentiment. Crimes are simply extreme cases, within a moral community in which there is a constant readiness to blame or criticize free riders, those who fail to do their fair share, or are exploitative or threatening. This is the essence of retribution: holding one another accountable. It is not peculiar to crime, but is manifested everywhere in society in our mutual accountability for meeting community expectations.

If accountability is an essential feature of community, and criminal punishment concerns its dramatic extreme, there are implications for social control. There is a continuity from mild moral offenses the law leaves alone to serious crimes. The morally important feature of this continuum traditionally has been a delinquent's degree of disgrace. Fining, jailing, or executing someone need have nothing to do with punishment. Punishment is essentially a moral notion and has no necessary connection with violence or deprivation of freedom. It is, ideally, putting someone in disgrace and the eliciting of repentance. If we just fine someone, we merely impose a cost; if we just take him out in the yard and shoot him, we are simply disposing of a problem, as we do with a disease or a mad dog. But, if we try to elicit a confession and contrition, we are seeking a moral transaction with a fellow human being with whom we share at least some principles. This may be one reason why regimes whose moral foundations are the most dubious often try the hardest to elicit confessions from prisoners and thus give themselves the appearance of moral legitimacy.

It might be objected that harm, in the form of fines, jail, or execution, is essential to punishment. This is right *if* one also accepts disgrace and ostracism (and the fear of them) as forms of harm. They can, after all, be feared and hated more than a fine or beating. Nonetheless, it is a mistake simply to put disgrace and ostracism on a list of harms. It is the same mistake J. S. Mill made in treating the "internal sanction" of conscience as merely a pain, like toothache; for disgrace and the threat of ostracism, in addition to being harms, are moral notions essential to punishment and are what make punishment a moral transaction. You can harm a mad dog but you cannot punish it because you cannot disgrace it, and you cannot disgrace it because it does not share your moral community.

It follows that it is impossible to punish some people, for if they are completely alien or sufficiently alienated, they cannot be disgraced

and they welcome rather than fear ostracism. An Amazonian tribe can harm me but it cannot punish me for violating its taboos. I must to some degree belong to a group, recognize it as one of my moral communities, before it can punish me. Hence, completely alienated criminals can be disposed of or quarantined, but they can be punished in at most a secondary sense. That is, as alienated rather than alien, they belong to my moral community and therefore ought to be susceptible to disgrace and fear of separation, but they are not and yet we impose the rituals of punishment on them anyway because we still claim these criminals as ours.

In tribal settings, fines or beatings are often more important for their symbolic role in a ritual of moral censure and disgrace than for the material harm they do. Harsh words, hard looks, temporary ostracism, and disintéressement are punishment when the object of this treatment has an adequately developed social identity. This is what punishment always has been in tribal society, in families, and in close-knit modern societies. To the extent that ritual censure is replaced by simply damaging someone—by fining, confining, or killing—to accomplish some future (utilitarian) goal, punishment has become another thing, though the name may stay the same. Hence, where dangerousness does not require jail, one should think first of punishments that put the potential delinquent at risk, in his local community, for disgrace and making restitution.

Putting crime in a completely separate category from misdeeds we blame without the aid of the law leads us to demoralize crime, which is more clearly a mistake than the decriminalization of immorality: when criminals are not seen simply as great violators of the mutual moral expectations in our daily social intercourse, but instead are viewed as mere harmful agents, they are removed from our moral community. They are removed because of a philosophical mistake that alienates them more than they were before. Retribution consequently is intrinsic to the cohesion of a moral community. This does not imply we should be constantly censorious or rant at one another; it is that we risk disapproval, or withdrawal of trust or privileges, when we disappoint normal expectations of cooperation, loyalty, and doing our fair share

There is a problem beyond the scope of this essay about whether "moral communities," as I have defined them, can ground moral rights or duties. It may be thought that my moral communities can be only "relative" or subjective grounds, since they are bounded geographically or, alternatively, are bounded by the idea of my own kind under a certain description. And hence it may be argued that moral duties can legitimately arise only within a single, objective moral

community—perhaps Kant's community of rational beings. One reply is that I take the notion of "moral duty" to be univocal in various moral communities, and, if there is a single core meaning, derivable from an objective conception of a global moral community, this is what I mean by 'moral duty.'

But I doubt the existence of a single, overriding moral community on which all conceptions of moral duty are dependent. More plausibly, every moral community is group egoistic, grounding rights and obligations that are both prima facie and relative to the good of one's group (under some description). It is possible that even Kantian and utilitarian conceptions of the domain of morality are "tribal" in the sense of being expansive conceptions of one's kind. It is also quite possible, however, that there exist moral ideals such as rationality and happiness which are independent of the notion of a moral community as I have defined it—that is, ideals describable without the use of egocentric particulars and hence independent of any conception of the good of my community, my species, my fellow rational beings, and so on. If so, what protects the good of one's community conceivably could clash with these ideals. I discuss this problem at length elsewhere, but cannot pursue it further here.

Some retributivists argue that punishing criminals is a form of communication which tells the criminal how wrong the act is, or hopes to show him it is wrong, or tells him that society condemns it. These views have been called teleological retributivism, however contradictory that label may sound. Although I agree that one function of punishment is to send a message to the criminal, I think the main message is something else: on the communitarian view I am defending, what a moral community communicates to criminals by punishing them is that they still belong to the community, they are still members. For it is our own kind that we hold responsible; we do not punish wild animals, bacteria, or total aliens, but instead avoid or dispose of them. Institutionalized retributive punishment is a ritual act that includes the criminal in the ritual, and a society does not include outsiders in its rituals. Although this is not its only or even its main purpose, retributive punishment (excepting execution and its analog in tribal societies, banishment) aims at a criminal's contrition and reintegration into the society.

Robert Nozick³ has suggested that the function of retributive punishment is to reconnect the wrongdoer with correct values; it is a

² I examine the relations between group egoistic and impersonal morality in Normative Behavior (Washington: University Press of America, 1983), and in The Non-Suicidal Society (Bloomington: Indiana UP, 1986).

³ Philosophical Explanations (Cambridge: Harvard, 1981), p. 374.

way-of-last-resort for correct values to have an effect in our lives. Since wrongness qua wrongness has no causal effect and the right values failed to motivate the criminal, we make, via punishment, the connection in his life with values that he failed to make by living morally. I think Nozick is right in affirming the importance of "reconnecting" the criminal with the moral norms of the community. But his view, like the communication theories, also appears to justify retribution in terms of what it does for the wrongdoer instead of what it does for us.

I suggest, on the contrary, that a moral community exacts retribution for its own good and not primarily to inform, connect, cure, use, or to send any kind of message to the criminal. All of the above are useful and at times desirable, but the essence of retributive punishment lies elsewhere. We would not punish Hitler, Josef Mengele, or a brutal rapist-slayer, primarily in order to rectify their relation to the universe. We would kill or imprison them because of what they did to us, and have no self-respecting alternative.

Punishing serious crime is one of the rituals that contributes to a society's identity as a living moral community, and simultaneously, as Susan Jacoby⁴ has said, denies promiscuity to the retributive urge. Like a bear dance or a funeral, it contributes to a society's identity as a living moral community. In the words of the criminologist Martin Levin,⁵

Our penal and judicial systems serve other goals than lowering the rate of recidivism. And the tension among these goals cannot be resolved on utilitarian grounds; one reason is that the punishment of criminals is, in part, a symbolic activity that expresses our ultimate values (*ibid.*, p. 103).

The importance of retribution is that, without taking proportional retribution in grave cases, a society dishonors itself, and undermines public confidence that the society takes itself and its values seriously. If Israel, for example, forgave the Nazis it captured and convicted, it would compromise the dignity and honor of the Israeli state. The strength of our indignation and demand for retribution depends both on the heinousness of the act and on the extent to which the delinquent is thought of as one of our own. This is why we do not view tribal head-hunting the same way we view ordinary murder by a member of our own community, and why genocide by a modern industrial nation is so executed.

These reflections are bound to make some people uneasy. Affirming the importance of ritual suggests an organic theory of society.

⁵ "Crime and Punishment and Social Science," The Public Interest (Spring 1972).

⁴ Wild Justice: The Evolution of Revenge (New York: Harper & Row, 1983), p. ⁵ "Cri-

And the organic theory has been part of a fascist conception of the state according to which persons are of value only as parts of a larger whole, dispensible and of no intrinsic value as individuals. Given the connotations of the "organic theory," it is better to speak of a communitarian theory of society, which puts one in the company of Tocqueville rather than Mussolini. It is, I think, a mistake to abandon criticism of radical individualism because of what Charles Taylor⁶ called the "totalitarian menace," a phrase he used when he was criticizing the "negative theory" of freedom. Ritual may remind some of Hitler's Nürnberg rallies, but it remains the most significant way in which people, world-wide, distinguish the act of a community from that of an individual. Marriages, convocations, commencements, celebrations, inaugurations, school recognition ceremonies, holiday observances, and countless other rituals help us take common goods seriously, maintain social identities, and diminish alienation, and it would be absurd to abandon or denigrate them because Hitler had Nürnberg rallies.

Second, moral communities as I construe them are multiple, nested, overlapping, and generate only prima facie obligations for their members. Any obligation to one's community (whether family, nation, or species) can be overridden by stronger obligations to other moral communities, or by moral principles independent of communities. The idea of a moral community imposing obligations that swallow up the individual is more tempting if one thinks there is only one genuine moral community.

Finally, I think the evidence is overwhelming that people need social identities—essences, as it were—and that otherwise they feel isolated, alienated, and without significance. Unless people noninstrumentally value something other than themselves, they will find it next to impossible to value themselves. This is what it means to say we are innately social animals and it is the basis for criticism of radical individualism. But it is one thing to say people cannot flourish without social identities and a sense of possession toward their societies, and another to say they are mere parts of society like bricks in a wall. If I cannot flourish without belonging to moral communities, it does not follow that I have no value as an individual and can be sacrificed for the community; indeed, one of the pre-eminent values of my community may be the value of individuals. Again, it would be absurd to abandon communitarian social institutions and become alienated because Mussolini viewed citizens as dispensible parts of an aggressive nation state.

⁶ "What's Wrong with Negative Liberty," in *The Idea of Liberty*, Alan Ryan, ed., (New York: Oxford, 1979).

II. SANITIZED REVENGE

Can revenge be kept out of this? Can we not show that retributive justice is the moral idea that some criminals deserve to be punished, and that this is a distinct moral concept with no necessary connection with revenge? Can we not affirm a normative retributive premise that makes no reference to desires or to moral psychology? C. W. K. Mundle, for example, offers such a premise. He says, "The fact that a person has committed a moral offense provides a sufficient reason for his being made to suffer" (ibid., p. 221). But, if someone simply says this claim is false, there is nothing more to be said by another who thinks it is true. And this is not its only difficulty. It affirms; but it does not explain anything about retribution's universality, origin, social function, or relation to retaliation. A philosophical concept of retributive desert, separated from both vengeance and utility, is a bloodless notion which cannot begin to account for the vehemence with which one seeks to punish the acts of Nazis and brutal criminals. Utility can tell us, roughly, how much to punish a kind of crime to reduce its future occurrence to an acceptable level. The idea of revenge makes understandable (whether or not we reject it on moral grounds) people's insistence on harming certain criminals. But retribution, distinct from utility and revenge, is a philosopher's phantom, getting what sense it has from its secret association with the idea of vengeance.

Can vengeance be a moral category, or is it inherently barbaric, destroying the morality of what partakes of it? As we have seen, we cannot have a moral community unless its members are personally accountable for what they do; personal accountability makes no sense unless it implies that transgressors deserve punishment—that is, they are owed retribution; and I shall add now that there is no doubt that retribution is revenge, both historically and conceptually. I do not believe that the logic of this chain can be broken—from moral community, to accountability, to retribution, to revenge. Neither do I believe it needs to be.

The solution lies in seeing that judicial retribution is not mere revenge but revenge that warrants its new name by satisfying certain social conditions. In simple revenge or retaliation the victims (or their relatives) set and carry out the punishment. Such retaliation is unpredictable from case to case, sometimes gets the wrong person, often is ineffective because the victim is not strong enough, and usually is harsher than nonvictims would think is fitting. Personal retaliation also has no built-in mechanism for termination, the result

⁷ "Punishment and Desert," *Philosophical Quarterly*, IV, 16 (1954). CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

often being an endless feud. This situation is not totally unlike what John Locke conceives punishment to be in the state of nature.

If revenge is the source of the passion and emotion behind our insistence on just punishment of criminals, it is impossible to eliminate it; we can only cleanse it, civilize it. As Jacoby says,

The taboo attached to revenge in our culture today is not unlike the illegitimate aura associated with sex in the Victorian world. The personal and social price we pay for the pretense that revenge and justice have nothing to do with each other is as high as the one paid by the Victorians for their conviction that lust was totally alien to the marital love sanctioned by church and state (op. cit., p. 12).

I suggest that when certain empirical conditions are met, retaliation, "getting even," turns into retributive justice, a moral concept. I call this *justice as sanitized revenge*. The conditions are that punishment:

- (a) is applied by officials who are not friends or relatives of the victim or defendant;
- (b) is applied consistently for similar cases, and hence is predictable;
- (c) is applied in accord with publically promulgated procedures and penalties.
- (d) is decided and pronounced in a context of ritual and ceremony, thus conveying that a community and not just an individual is speaking.
- (e) is decided after due deliberation and not in the heat of passion.

These conditions require refinement and perhaps additional ones will need to be added. When enough such conditions hold, we have retributive punishment, a moral act, and not mere revenge. I do not suggest that these are necessary and sufficient conditions. "Retributive justice" is an open or cluster concept, the relative weights of the conditions that turn revenge into retributive justice depending on their degree, how they are combined, and the particular circumstances. Neither does this conception of retribution yield a formula for determining what punishment one should believe "fits" a particular crime. We have general feelings for what is "too much" and "too little" punishment, and in between we depend on a combination of reciprocity, the degree to which the crime is feared and hated, predicted utility, and unexplained (perhaps unexplainable) feelings of appropriateness.

A law that is imposed unequally for equal crimes, or imposed by vigilantes instead of by agents of social institutions, is rightly said to be unfair. Punishment by society protects you from your victim's enraged relatives, the fear of which could last forever, and substitutes solemnity and ceremony for their rage and unpredictable violence. My idea is that nonmoral retaliation simply turns into retributive justice when enough of these empirically describable conditions CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

are satisfied; there is no special ingredient, metaphysical, moral, or theological, that must be added.

Revenge "turns into retribution" in that where it would be true that I desired out of revenge that C be harmed, if we change the case just by adding the conditions, then it is true that I believe that C deserves to be punished. There are certain other requirements such as my believing the conditions are met, but I need not believe these conditions turn revenge into retribution. People can be mistaken about whether they seek revenge or retribution, just as they can be mistaken about whether they have a moral or a nonmoral belief.

When revenge is expressed through institutional ritual and ceremony, the actions of individuals are transformed into actions of a community; the private act of an individual becomes, through ritualization, the moral act of a collectivity. The indignation of the community is expressed in solemnity, black robes, and ritual, which is a transformation and sublimation of anger and invective. If all of this is so, revenge is not eliminated and replaced by something totally different, but only civilized, "sanitized." In distinguishing retributive justice from revenge, we do not take the revenge out of retribution—this cannot be done—but show how circumscribing and institutionalizing retaliation turns it into a moral category, warranting a new name. The socialization of revenge transforms it into justice, but possibly false justice, a problem we must address.

The anthropologist Christoph von Fürer-Haimendorf⁸ hypothesized the evolution of criminal justice from personal retaliation. There is a moral development from private to public intervention as a society becomes larger, more stable, and, I would add, develops a sense of a common way of life which needs to be protected. Among the primitive Daflas, a non-Hindu tribe in India, Hobbes's war of all-against-all began at the door of each long house, and "justice" was no more than retaliatory raids by relatives. The more complex society of the Gonds, another non-Hindu group in India, created a sense of social identity among as many as forty or fifty villages. The size of Gond society ruled out personal retaliation, Gond public opinion demanding institutionalized procedures because many people now felt they had a stake in the quarrels of strangers; they wanted punishments to have authority, predictability, and to make sense. The size and complexity of Gond society constituted a threshold in the evolution of criminal justice out of personal retaliation.

The substitution of the principle of retribution at the hands of kinsmen by a judicial process and the imposition of legal sanctions by the community as a whole, marks a distinct stage in the development of moral and

⁸ Morals and Merit (Chicago: University Press, 1967), p. 105.

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

judicial concepts. . . . This development of a public sense of morality and the assertion on the part of society of a right to intervene in private disputes of individuals has nowhere been a sudden process (*ibid.*, p. 105).

III. THE PROBLEM OF JUSTIFICATION

One may object that it is unimportant whether people believe that someone deserves retribution; what is important is whether the belief is true. The two arguments I presented conclude that morally accepting retribution is necessary to a viable moral community, and that revenge, when certain empirical conditions hold, turns into retribution. They do not conclude that people are morally accountable or that they deserve retribution. Of course, if we agree that we ought to have viable moral communities, then it follows that we ought morally to accept retribution: but not even from this can one deduce that anyone deserves retribution. Thus, there are two normative claims one would like to establish. The first, 'We ought morally to accept retribution', is easy, requiring only the premise 'We ought to do what is necessary to flourish and achieve our full humanity'. It is not that I have the slightest idea how one would prove that premise, but only that few would dissent from it. The second normative claim, 'Some people deserve to be punished', is more difficult, and I know of no noncontroversial moral premise from which it follows.

What would follow if every claim that someone is morally accountable and deserves punishment is false, notwithstanding that holding people accountable in this sense is essential to the life of a moral community? I am not supposing that all retributive judgments are mistakes in the sense that everyone judged is mad or innocent, but that on philosophical (e.g., utilitarian or deterministic) grounds no one ever is morally accountable. It would seem to follow that we could not justify what is essential for a moral community, and therefore could not justify having moral communities as I have defined them. Following this out would require explaining moral communities without moral accountability, or arguing that we ought not live in moral communities, or that we should become extinct. Another possibility is that what is essential to the life of a moral community is holding people morally accountable, whether or not they really are morally accountable.

Where then do we stand? I have not argued for the claim, 'People sometimes are morally accountable and deserve punishment', but I have given what appear to be utilitarian reasons for the claim, 'We ought to hold people morally accountable and deserving of punishment'.

There is a level on which utility is, indeed must be, relevant to our appraisal of any system of social organization. Moral community and

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

personal accountability must have utility, relative to other ways people might live, or else both biological and cultural evolution would have favored those other ways. But this speaks to the belief in and not the truth of retribution, if the latter even makes sense. It is equally certain that utility cannot be the reason one accepts retributive punishment, for then it would not be retributive punishment.

If we say retribution has utility, we could mean that we have a better society if people falsely believe members of their moral community sometimes deserve to be punished. This is the "noble lie" view. It is, I believe, elitist in the bad sense; the very idea of philosophy seems absurd if philosophers must conclude that the best world is one in which the majority of people must believe what is false. Philosophy surely aims at the discovery of truth by everyone capable of doing so. Moreover, the "noble lie" view presupposes that retribution involves true or false moral propositions, and not just a set of cultural practices. If judgments of desert are not truths or falsehoods, we are not, as social philosophers, telling people to believe what we believe is false.

Alternatively, one might argue that retribution has utility just because people insist on it that much, and those deprived of seeing retribution done will suffer more than those who are punished. By this line of argument, anything people want badly enough on non-utilitarian grounds is justified on utilitarian grounds. It may be so, but, if it is so, it is not a way to defend retribution but a reductio of utilitarianism. The practices of racism, sexism, and ruthless competition, if desired so strongly that their deprivation creates more misery than their exercise, have utility; but they are unjust and wrong nonetheless.

I do not think it is intelligible to accept my arguments for holding people morally accountable, and at the same time suggest that all judgments of moral accountability are false in the literal sense that they all ought to be rejected. Ruling out the "noble lie" view, our options are (1) some retributive judgments are true, and proven so by arguments philosophers have not yet discovered; and (2) retributive judgments are justifiable independently of the issue of truth or falsity. I think the latter option is more plausible, for supposing all retributive judgments false and therefore to be rejected sounds unintelligible, given the acceptability of my arguments for holding people morally accountable; and it does so prior to our knowing whether or not we can demonstrate the truth of retributive judgments.

The idea of retribution I am explaining and defending is not like the old concept of retribution which played little or no role in events: when philosophers rejected retributive grounds for punishment, they usually just substituted utilitarian reasons for the same or very similar practices. Can we abandon accountability and retribution as I have explained them, and substitute utilitarian reasons for the same punishment policies and practices? Not in this case, if my argument is correct that holding people personally accountable, and not just punishing them, is necessary for the existence of viable moral communities. But holding people personally accountable is precisely what we abandon when we abandon retribution. In the one case we justify jailing robbers only to keep us safe, after we abandon the idea that robbers deserve jail. In the other case we hold them accountable because of what they did in the past, and not because it will make us safer in the future.

I would like to suggest that a retributive judgment is a move in the social practice of retributive punishment. I have argued that the social practice is essential for human social life as we know it. Hence, justifying retributive punishment is not like justifying the particular judgment 'Josef Mengele deserved to be punished', it is like justifying funerals, marriage, or bear dances. And such things can be justified. After the practice is justified, 'Josef Mengele ought to be punished' then may be justified within the practice, in terms of moral and legal reasoning about crimes and excuses.

This is not at all intended to be a contribution to the new fashions in "moral realism." Retributive judgments are true or false in at best an honorific sense; they are relative to practices, and social practices are called good because social animals such as we require them for cooperative living. The whole theory comes together more smoothly if some version of nondescriptivism is true. We do not have to worry about the objection that retributive judgments are false if they are neither true nor false. As for the judgment that we should preserve moral communities, I have argued elsewhere that fundamental moral judgments are expressions of basic wants that have causes but not reasons; and that these wants turn into moral beliefs in the presence of criteria ("marks of the moral"), analogous to those I have here argued turn revenge into retribution.

Holding people morally accountable and punishing and rewarding them is best understood as social ritual such as bear dances or marriage. It makes no sense to say that marriages or bear dances are false, but does make sense to say we ought not to have them. On this view, a mistaken retributive judgment is a defect in a different category from that where we give evidence that a social practice is not utile. It is more like doing the bear dance incorrectly.

ANDREW OLDENQUIST

Ohio State University

⁹ Normative Behavior, ch. 3.

MOTIVE AND JUSTIFICATION*

ORALITY presents a deep problem about the relation between motives and justifications. It is that the rational motives a person has to act morally do not seem to be as strong as the obligation morality imposes. For (it seems) the obligation must outweigh any opposing considerations about how to live; and self-interest is the only motive to which a reasonable person can give such absolute weight in planning her life; yet it does not recommend absolute obedience to morality.

When the ancient philosophers considered this problem, they all started from the belief that only self-interest can be a reasonable and compelling motive for a person's life. They then needed to argue that a moral life is a good for its agent. This is a hard view. It allows two possibilities. One is Plato's idea that virtue is intrinsically enjoyable, in the way an exercise of skill can be, so that it belongs to the content of happiness. Epicurus chose the other—virtue has beneficial outcomes, in the way good sense and self-control do, and so is a means to happiness. Neither idea seems plausible about real life.

The starting point for most post-Kantians has been that a morally good life is not a self-interested one. So, they have needed to hold that virtue can have a different motive which is both reasonable and compelling. This is an equally hard view. No end seems to weigh enough. For instance, although another person's good is a sensible aim, it will not withstand the moral crunch. You have a reason to let me go first in the grocery line if you have two carts full and I have a box of fudgesicles. This does not mean you will see the point of letting me go first into the lifeboat, for the last seat, even if I got in line ahead of you.

I shall argue that the ancients chose the right point of departure. It is intuitively certain, the alternatives to it fail, and the difficulties into which it forces us are not hopeless.

A prerequisite to defending their viewpoint is a more exact formulation of the problem itself. This requires definitions of the two operative terms and a statement of the conditions something must satisfy in order to be a motive to live morally.

By a motive I mean "that on account of which a person acts." It is a psychological term, in the same family as 'belief'. It names a desire or

^{*} The Northwest Institute for Advanced Study gave me a grant to work on the present paper over the Summer, 1987.

⁰⁰²²⁻³⁶²X/88/8509/0479\$02.10 © 1988 The Journal of Philosophy, Inc.

an end or anything to which one of these transfers its energy. A justification is a "reason showing that an action is right or ought to be done." Like 'proof', it expresses a normative idea. Since 'right' and 'ought' bear several senses, a moral justification is only one kind among others. Further, I shall use reason for acting as a neutral phrase expressing either concept. The two concepts are obviously distinct, since a person can have a desire to do something without thinking that she ought (in any sense) to do it. On the other hand, they are not necessarily exclusive, since the belief that an action is right for her to do might itself influence someone, as Kant thought about moral rightness.

There are three requirements a motive must satisfy in order to be one for morality.

(M1) It must be reasonable. It is so if a person is entitled to think, either self-evidently or from good evidence, that its attainment will be satisfying. In turn, a purpose is satisfying if the agent sees its point, if he approves of having acted for it, and if he can weight his approval consistently with his other purposes. Let me put this shortly by saying that a motive is reasonable if its attainment is likely to receive the agent's reflective approval.

This does not mean that it will further his good. I have used a definition of 'satisfying' that is not restricted to what conduces to one's own happiness. One can endorse purposes that do not benefit oneself. Both sympathy and revenge (to use Bishop Butler's examples) often have satisfying results. At the same time, the definition is not empty. It rules out wanting a transparent inconsistency or wanting an action for a feature one should know it does not have. It has some consequences beyond these minimal formalities. Meaningless activity for its own sake is not sensible, since an agent must see a point in what she does. Because it is formulated in terms of "approval," the definition makes self-respect essential—if someone approves of achieving a certain end, she cannot disapprove of herself for achieving it. So, every reasonable purpose must be consistent with self-respect, and, so, the desire for it must be also reasonable.

'Reflective approval' is not the same as 'justification'. The first marks a sense in which something is satisfying; the second is about rightness; and neither 'satisfying' nor 'right' includes the other in its meaning.

The need for (M1) comes from observing that people have motives that they would abandon if they were clearheaded. One person unthinkingly imitates the quickness of his father's temper, although he destroys friendships he values. Another observes religious ceremonies whose meaning she has lost. Both carry over from childhood

Digitized by Arya Samaj Foundation Chennai and eGangotri

pointless habits. Moral behavior itself should have a better motive than arbitrary conditioning. So, (M1): it must make sense to see an ethical life as satisfying.

(M2) A moral purpose must extend as widely as moral obligation, in the sense that it must motivate any action that morality requires.

Although I say 'It must . . .', nothing important attaches to the singular pronoun. A combination of aims rather than a single one will work. For instance, a sense of duty might lie behind what Adam Smith calls the awesome virtues of self-command, and sympathy behind the amiable ones that are directed to another's good. In any case, whether the motive is one or many, the idea (M2) expresses is that there cannot be a leftover corner of obligation which has no motive.

Since the paradigm moral decision is about how to live, anything satisfying (M2) must apply to a person's choice of a kind of life. It must be a reason for choosing a moral one over an immoral one.

(M3) Its weight as a motive must match the strength morality has as a justification. For instance, if a moral reason for doing an action always overrides nonmoral ones against doing it, then a moral incentive must outweigh ones to act immorally: if duty is categorical, its motive must be conclusive. Otherwise, although there might be a reason to act morally, it would not be a strong enough one.

What is the strength of moral justification? Kant's view is that it is absolute, but this is both more than the problem needs and more than can be assumed without argument. If it were true, all ethical rules would hold no matter what. They would impose equal duties. This is not the view of real life, which gives greater weight to some rules than to others. Sexual prohibitions form a large part of ordinary morality, but most people do not think they are absolute. On the other hand, a central idea of conventional thought is that one should be a decent sort of person. This means roughly that one should not use anyone and should value human beings over nonhuman things. Most people do think that this command holds absolutely; or that it can be overridden only as a remote possibility, in conditions which they cannot easily imagine. This is what I assume, too: the basic moral obligation is to lead a life of human decency, and it is as serious as any concern can be. Particular moral rules might not (or might) have as much importance.

(M2) and (M3) are related conditions: they specify that an ethical purpose must match the extension and the depth of ethical justification. (M1) asserts that morality is a rational enterprise. The problem of motive and justification is to find an aim that fits all three conditions. I shall call anything that does a moral motive. It is possible that

several do, but the problem is to find even one. In terms of these conditions, the starting point of the Greeks can be expressed: only self-interest can meet (M1) and (M3). The post-Kantian one is: self-interest cannot meet (M2).

I am going to defend the ancients' choice of a starting point. This is less than a theory of moral motivation. Its main defense must consist in countering arguments that its appearance of self-evidence is only an appearance. Still, there is a reductio for it, as there often is for beliefs that are near first principles, and I give this in section III. The longer project is to defeat alternative candidates for (M1) and (M3). Self-respect, a sense of duty, and altruism are the ones that have current followers. I discuss them in section IV. (M2) cannot be left entirely offstage. For I assume that there is a motive to live morally. Some end must fit all of (M1)–(M3). So, the apparent impossibility of self-interest satisfying (M2) exerts pressure backwards on the claim that it alone satisfies (M1) and (M3). I discuss this problem in section V. I describe a viewpoint from which it is defensible that self-interest fits (M2), but I do not take the further step of showing that the viewpoint is true.

II. THE SELF-INTEREST THEORY

The form of the ancient position to which I shall refer from now on is:

(SI) The only reasonable, conclusive motive for choosing a certain kind of life is that it is a good for oneself.

'Reasonable' comes from (M1), 'conclusive' from (M3), and 'for choosing a certain kind of life' from both (M2) and (M3). By virtue of the last quoted clause, SI says that self-interest has the same scope as morality, which is a necessary condition for (M2), but it does not say that it has the same results within that scope. It does not say:

- (D) A life of human decency is always a good for the person who leads it.
- (SI) and (D) together are the bases of a full theory of moral motivation.

Some further explanatory remarks about what (SI) means are in order, so that it is not open to attack for what it does not say. The first set are about the concept of a "good for oneself" and afterwards about a "kind of life."

One difference between interest and passion is that the first is directed to a good and the second is not necessarily. Sexual desire has its moving force independently of the agent's approval. The word 'good' in (SI) marks this difference.

A good for oneself can be also a good for others. The identity can even be analytic. It is, for instance, for certain public goods like equal protection under the law, from the meaning of 'equal'. (SI) is consistent with the idea that the most important goods are public ones. So, it does not imply that the welfare of others is not a reasonable, conclusive purpose for one's life. All it implies is that what is not one's own good is an unreasonable purpose.

It is an open question whether (SI) is a maximizing demand. The argument that it is goes: "(SI) says that one should not choose a kind of life so far as it is bad or indifferent. If someone chooses less than the best possible life, he takes a loss. If he takes a loss, he chooses a comparative evil. So. . . ." One thing that is not obvious in this argument is that the best possible life is ever an object of choice, or that lives can be ordered by distance from a maximum. There is no sense in the idea that a chef should plan the best possible dinner. Even if the ordering is possible, there is a doubt that to choose less is to choose a comparative evil. Suppose I can steal \$30,000 or \$40,000 from the safe. \$30,000 strikes me as enough and I take it. Something can be sufficiently good to be the object of rational choice without necessarily being the best possible. In both (SI) and (D), 'good' is to be taken in this sense as "sufficiently good."

The differences between interest and passion, public and private goods, sufficiency and maximization, are all scholia on the concept of "good for oneself."

I need also to distinguish a policy for a *kind of life* from the purpose of an *action*. The first is for one's actions taken collectively rather than distributively. It is about individual actions only so far as they fit into an overall pattern. Now, the fact that a set of actions is done from a certain motive does not imply that each action is so motivated. For an individual move can have a different purpose from the strategy into which it fits, as tactical retreat has a place in a campaign to win territory. In the same way, a self-interested life can include generous actions. And it must: people who look out for themselves in everything and always want to be first in line are grubby like moles. So their lives must look even from their centers.

I am making a point that is structurally the same as the one Leibniz made about the best possible world. That is not a world in which each monad is the best possible, but it is the best compossible set of monads. Leibniz said this to defend the consistency of the views that God chooses the best and that many things are yet worse than they need to be. For God chooses the best compossible set, and many things are worse than they need to be when considered individually.

The same is true here: a life that includes some self-denying actions will be better as a life than one that does not. Self-interest requires a pattern, some elements of which are not self-interested.¹

(SI) entails no theory about the relation between time and a good life. A person does not plan her life from a position overlooking it, as the demiurge does the universe. She does not have the view that a utilitarian calculator is imagined to have of society. So, there is no similarity to ground the idea that, just as utilitarianism is indifferent to who has a satisfaction, so (SI) is indifferent to when satisfactions occur.

Apart from that nonexistent similarity, it is still a commonplace that choosing a good overall life requires equal concern for all parts of one's lifetime. The idea comes about from reflection on the fact that passion can move a person to his own harm. Sometimes (not always, since this does not fit revenge) passion makes immediacy seem to be a larger good than it is. The view that warps in the opposite direction is that temporal location is not a good at all. So, one's concern for all phases of life should be equal. I do not know if this is true. It is not built into (SI), which requires only that a person choose a life for being a real and not merely an apparent good, and which does not have any theory about what separates the two. As far as I can tell, all important truths about the relation between time and a good life must be learned from experience.

The relations between a kind of life and individual actions and between a good life overall and time are supposed to clarify the concept of a "kind of life." I hope also that the concept of "good for oneself" is in order, as well as the relation of (SI) to the conditions defining a "moral motive," and that all these combine to make (SI) itself a clear sentence.

III. AN ARGUMENT FOR (SI)

Although there are arguments against particular alternatives to it, (SI) is so close to self-evidence, that it can have only the kind of intuitive reasons sometimes used to bolster the already obvious.

I noted before that self-interest is not the only reasonable motive.

¹ Derek Parfit [Reasons and Persons (New York: Oxford, 1984), pp. 5–12] gives this argument to show that (SI) is "indirectly self-defeating": (a) If one wants one's life to go as well as possible, then one does not want to perform any self-denying actions. (b) If one does not perform any self-denying actions, one's life will be worse than if one did. (c) So, if one tries to make one's life go as well as possible, one's life will be worse than if one did not. (c) is the claim that (SI) is indirectly self-defeating. Parfit explains with some ado that it does not imply that (SI) is false. Since (a) is false, there is no reason to accept (c) nor to let pass the idea that (SI) is self-defeating, with its sound of a forgiven fault.

It is one, however. For self-respect is reasonable, and it requires a person to see her own good as worth pursuing, in the absence of special reasons not to do so. Self-interest is always a sensible purpose. This does not mean that it is compelling or overriding. A person can reflectively approve of self-sacrifice, if it has a point consistent both with her other purposes and with self-respect. Now, I think these conditions hold only for particular actions. They do not fit one's choice of a kind of life.

Consider why it is reasonable that some actions should not be self-interested. Some are throwaways—they have a negligible effect on one's own good. If a person could use these to further someone else's good, he would be mean-spirited not to do so. The charity of the well-off and the courtesies civilized people show each other are like this. Other actions, like generosity to one's friends, must be selfless to fit into the pattern of a good life. More generally, to be happy one must have a certain largeness of view—a detachment from oneself and an empathy with everywoman.

Yet the considerations that make selfless actions reasonable do not apply to a person's life. It is not a throwaway and it does not fit into anything further. Suppose, then, that someone dedicates himself to a purpose that is indifferent to his own good. There are only two possibilities. The first is that his end is not a good for anything. Then he lives absurdly, as if he were counting bricks in a wall, which is also a selfless action. The second is that he adopts someone else's good as an end, to which his own good becomes a means. Then he has subordinated his life and his welfare to the welfare of others. So, he is used, even if he is used voluntarily. It is not reasonable to live absurdly or to be used.

The first possibility covers a Kantian candidate for a motive. A sense of duty is an empty purpose if it is not directed to any good. Now, Kant does explain its point, and I shall discuss this in section IV. If he had no explanation, if duty were blind, then being dutiful would show no more sense than being compulsively tidy. The second possibility is based on Kantian considerations. If we are drawn to the idea that the moral law is not to use anyone, including oneself, then altruism cannot be a reasonable motive for one's life. To be altruistic is to act for another's good in indifference to one's own. There is nothing awkward about being indifferent to one's own good in particular actions, since these are often unimportant or have a place in a good life. There is nothing awkward about living for others, when it is consistent with living for oneself. But to live selflessly for the good of others describes servility. Self-respect is a limit one's own good puts on altruism.

The possibilities other than self-interest are that a person lives for something which is not a good at all or for something which is not her good. To reject these two as absurd or subservient is to return to the original position, that a reasonable life must be a good for the person who lives it.

IV. ARGUMENTS AGAINST ALTERNATIVES TO (SI)

The main arguments for (SI) are against specific alternatives to it. An alternative is a motive put forward as fitting (M1) and (M3). [The ones I consider are intended also to fit (M2) and, so, to be moral motives. This matters only indirectly to the defense of (SI).] The suggestions now on the philosophical board are duty, self-respect, and altruism. These are not random choices from an indefinitely long list. On a certain conception of ethics, there will be only two general classes of possible moral motives other than self-interest. The form of morality suggests ones deriving from the concept of principle. Duty and self-respect are typical members of this group. The matter of morality is consideration of others for their own sake and, so, suggests motives of sympathy—for instance, altruism. There are an indefinite number of members within each class, but I do not think the argument turns greatly on differences among them. The live issue is whether the basic idea of either group will work.

IV.A. Self-respect and a sense of duty. Kant's (and later H. A. Prichard's) view is that morality is itself a motive.² The fact that a person has a conclusive justification to do an action itself motivates her to do it. Nothing further like self-interest is necessary—nor sufficient for that matter, since no one is morally praiseworthy for wanting to benefit herself.

Kant has two explanations of how duty can be an incentive. He has first a theory about how a cognition can cause an emotion (in the present case, a feeling of reverence for the moral law), which then motivates. Ignore the quaint psychology of this. It is not the right

Prichard's views are in "Does Moral Philosophy Rest on a Mistake?" in *Readings in Ethical Theory*, 2nd ed., Wilfrid Sellars and John Hospers, eds. (New York: Appleton-Century-Crofts, 1970), pp. 83–96.

² Kant says in, among other places, his *Critique of Practical Reason* [Lewis White Beck, trans. (Indianapolis: Bobbs-Merrill, 1956), pp. 74/5 (Ak. II: 71/2)] that obligation is a motive. His explanation of how a cognition causes an emotion is on pp. 76-79 (Ak. II: 73-76). That freedom is an incentive for morality is from Grounding for the Metaphysics of Morals, James W. Ellington, trans. (Indianapoin: Hackett, 1981), pp. 51/2 (Ak. IV: 449-450). Kant's views about self-respect are 35/6, 39-41; Ak. IV.428/9, 433-436). Kant's first explanation is in terms of respect the law; his second is in terms of personal respect. The Kantian relation between as bound by the law.

kind of explanation. For a causal theory about how one comes to feel motivated does not show that one's purposes are reasonable. It does not show that it makes sense to feel reverence for the law and, so, does not establish that (M1) is met.

Prichard's view has the same difficulty. He believed that a sense of duty influences. It does, of course, and it can be strong enough to fit (M3). The question remains whether it is rational or not. For it might be only the residue of childhood training and, like belief in a fundamentalist religion, something to be outgrown with college and enlightenment. Prichard walked away from the problem of showing that a sense of duty satisfies (M1).

Kant does better with his second explanation. He says that a concept of oneself as free, or self-respect, is a motive. (It is not clear whether he intends these to be two motives or one. Unless his words have a special twist, the idea of oneself as free is a necessary part of self-respect. So I take him to be speaking of only one motive. He is giving two accounts of how it moves—one turns on the meaning of 'free' and the other on that of 'respect'.) Now, it is reasonable to want self-respect, and the desire cuts deeply enough to match the categoricity of morality. Further, both of Kant's reasons for saying that it extends as widely as morality have great charm. First: to respect someone (including oneself) is to grant her an unconditional right to act for her own reasonable purposes. A right is the inverse of a duty; an unconditional right of a moral duty. Self-respect, then, derives from the idea that one is bound by moral duty. Second: to act from the moral law is to act on the basis of justifications. To act so is to remove the unconscious causation of present actions by past events. It is, therefore, to attain rational self-control or freedom, similar to that which psychoanalysis can reach.

If Kant is right about all this, it does not follow that moral action has a motive other than self-interest. I do not see how he thought it might follow. For self-respect is an obvious part of anyone's welfare. Surely, the fact that everyone requires for his happiness a sense of worth is too visible to be ignored or to need defense. So, Kant's best argument, if successful, explains how moral action is essential to an agent's welfare

It is a reminder that self-interest is not directed solely to what the Stoics called "goods of fortune." Unreflective people occasionally identify their welfare with a standard of living. It consists in owning packaged goods and services. They see the "spiritual goods" of religion in the same way, as things to be received from outside a person. Yet everyone knows on reflection that interest is directed also to active states of the self. People want both self-respect and self-es-

teem. This is not news to any philosophical tradition, but Kant must have lost sight of it when he presented his view as an alternative to a self-interest theory. He was, rather, identifying a psychological attitude that can be a common end of both interest and morality. He was

supporting a self-interest theory.

IV.B. Thomas Nagel on altruism. Until recently, the problem of motive and justification had been quiescent in twentieth-century analytic thought, because theorists wrote about ethics primarily as a system of justifications. They simply did not stew about how it motivates. Nagel and Philippa Foot revived the discussion.3 Nagel argued that altruism is a reasonable motive that is internal to morality. Foot's view was that duty motivates only if it serves purposes external to it. They brought forward two issues at once. One is between the ancients and Nagel: Is the moral motive self-interest? The other is between him and Foot: Is it an intrinsic feature of right action, or is it a further purpose that morality serves? My concern is only with the first question.4 (The two are independent. Plato and Epicurus agree on the first and take opposite sides on the second.)

In trying to set the starting point for an answer to that question, I

³ Nagel, The Possibility of Altruism (New York: Oxford, 1970); and Foot, "Morality as a System of Hypothetical Imperatives," The Philosophical Review, LXXXI,

3 (July 1972): 305-316.

(M1)–(M3) are also criteria of adequacy for the solution of the second issue. For instance, both Epicurus and Hobbes believed that no intrinsic aim can match the seriousness of morality. According to Epicurus, only pleasure (Hobbes: only self-interest) can be a dominant motive for a person's life; only such a motive has the gravity of morality; and it is extrinsic. A typical connection between morality and pleasure, Epicurus says, is that everyone who acts unjustly feels anxiety and, so, forfeits peace of mind. Peace of mind is both enjoyable in itself and greater than other pleasures. Just action is a necessary means to it. (Hobbes had similar thoughts: keeping contracts is an essential means to personal security.)

Foot and Williams note that aims intrinsic to morality are not always reasons for an agent. They are only when they connect with purposes a person already has. Unlike Epicurus, neither writer has a reductionist view about purposes. These can be anything that make sense to a person—pleasure or not, self-interest or not—but morality must fit with them if it is to make sense. Foot (op. cit.) argued that the intrinsic point of morality is no more transparent than that of propriety. Both obligate apart from external ends, but etiquette motivates only someone with a use for it. So also might morality. Bernard Williams ["Persons, Character and Morality," in Moral Luck (New York: Cambridge, 1981)] explained that a person's deep motivations are the projects that make his life meaningful and that ethical values become motivated by fitting in with these. It would not be reasonable for a person to forfeit the sense of his life to satisfy the abstract demands of duty.

For Epicurus and Hobbes: no internal moral aim can meet (M3). For Foot and Williams: none can meet (M1). So, morality must be motivated from outside. In these ways, both arguments reach their conclusions by first eliminating internalism. On the other hand, arguments for internalism are not reductios. Rather, they try first to identify a motive within morality, so that it automatically fits (M2), and then

have no reason to take issue with Nagel's line of thought. For, even if it is sound, it does not establish an alternative to (SI). Suppose that practical rationality requires that the good of another is often an immediate reason for me, and that the ethical demand that needs to be motivated is the demand that I act for another's good. Then, as a requirement of practical reason, altruism is a motive internal to morality. Let the conclusion be given. If altruism is a reasonable aim, it fits (M1); that it is internal is sufficient (but not necessary) for it to meet (M2); but neither claim implies that it is a conclusive motive. They are both irrelevant to (M3).

It is important to be clear about the difference between saying that a reason is *intrinsic* and saying that it is *categorical*. For these two represent a further conceptual difference beyond that between "motive" and "justification."

Kant defined 'categorical imperative' in two senses. He defined it first as an assertion that an action is good in itself. This is not equivalent to his second definition, that it says that something is to be done, no matter what. I shall use 'categorical' only under Kant's second definition. I shall also slip into talking as though I hold his view that all morality is categorical, although my position is weaker, because it is convenient to avoid repeated qualifications.

Something good in itself is desirable for its own sake rather than for its role in a wider project, as a person might learn French for his own amusement rather than for business. His motive is then intrinsic to his activity. This does not say anything about the importance of learning French compared to his other purposes. On the other hand, a categorical consideration overrides anything that conflicts with it. It has absolute weight relative to other reasons. Something can be intrinsic without being absolute: usually, nothing is categorical about

⁵ This is the argument Nagel gave in *The Possibility of Altruism*. He describes its strategy on pp. 13–17. He later dropped most of it, because he came to think his reasons were partly wrong but also unnecessary—the conclusion is evident on its own. He explains his change of mind in *The View from Nowhere* (New York: Oxford, 1986), pp. 160–162.

⁶ The first definition is at *Grounding*, p. 25 (Ak. IV.415); the second is on p. 29 (Ak. IV.420). Kant argues that the two are equivalent when he gives the second one: "The categorical imperative alone purports to be a practical law. . . The reason for this is that whatever is necessary merely in order to attain some arbitrary purpose can be regarded as in itself contingent, and the precept can always be ignored once the purpose is abandoned. Contrariwise, an unconditioned command does not leave the will free to choose the opposite at its own liking. Consequently, only such a command carries with it that necessity which is demanded from a law." His argument has two slips: first, what is necessary for a *further* purpose and what is necessary for an *arbitrary* purpose are not the same. Second, the "contrariwise" sentence does not follow anyway. There is no inference from 'If for a further purpose, then contingent' to 'If not for a further purpose, then necessary'.

a person's reason for acquiring a new language. He does it as a diversion rather than as a career or a duty. It is enjoyable in itself but carries little weight. So also, what is categorical is not always intrinsic, since a means necessary for an end must have the same weight as the end. If the latter is categorical, so also is the former. (This sounds odd because a means is the content of a hypothetical imperative, which is the opposite of a categorical one. Keep in mind, however, that they are opposites only under Kant's first definition.) Suppose I have promised to help you, and the only way is to give you a hundred dollars. Then I am equally obligated to give the money. I do it as a means to discharge my original obligation.

'No matter what' weights a reason relative to others; a reason 'good in itself' gets whatever weight it has from itself rather than from outside. So, that an end is internal to morality says nothing about its importance. It is not relevant to showing that the end satisfies (M3).

'Reasonable' also is not equivalent to 'categorical'. It has two senses—either "approved by reason," which is the one I used in (M1), or "required by reason," which is its sense in Nagel's conclusion. Obviously, something can be subject to reflective approval without being categorical. Suppose that, after studying law, someone decides to become a public defender. She chooses reasonably if the career is honest and provides a living and so on. This does not mean that she was categorically required to choose it nor that, having chosen it, it must outweigh all later options. For that matter, a person acts with rational approval and perhaps without self-concern in undertaking the preservation of a historic building. Yet he might be unreasonable if he let this aim override his other concerns.

What is required by reason has a better claim to being categorical, but it still does not have a valid one. (A purpose is required if it would be irrational not to have it.) For a person can be unreasonable in this sense in small matters, as someone who feared raisins would be, and as everyone often is in postponing irksome chores. It is a heavy-handed thought that irrationalities of this sort are to be avoided, no matter what. That an action is required by reason does not mean that it is categorically required.

There is a further point to be made about the claim that one is rationally required to have a certain motive. 'To have a motive' does not mean "to weigh it more than others." Suppose I am required to recognize another's good as my own purpose. This does not imply that I must recognize it as compelling or urgent. It means: another's good must have a place somewhere in my psychological economy. Not necessarily: it must have first place.

Just as 'intrinsic' is not the same as 'categorical,' neither is 'approved by reason' nor 'required by reason.' An approved policy is not a required one. A required one need not be required at all costs, and a motive one must have is still not one that must outweigh all others.

Nagel's argument does not show that a motive other than self-interest can approach the gravity of moral obligation. It is like Hume's remark that people have in them "a touch of the dove" which responds to virtue. Probably so, but a touch does not propel anyone to the effort of a morally decent life. Hume wrote as though a moral skeptic was someone completely unwarmed by sympathy. Yet a person is also a skeptic if she thinks that morality has only the appeal of a hobby and its rejection the irrationality of a mild phobia.

If successful, Nagel's argument would show that altruism fits (M1) and (M2), but it leaves (M3) to fend for itself. What can be saved from it, then, is the reasonableness of being moved by another's welfare. I noted in section II that this is consistent with (SI). Since morality does require one to act for another's good, one ethical

command is consistent with (SI).

IV.C. Parfit on altruism. Parfit improves Nagel's position.⁷ He claims that something other than self-interest can be reasonable as an ultimate aim for a person's life. This does meet (M3). He also expands the repertoire of arguments. One points out that a person can be both reasonable and non-self-interested if he has an achievement as an ultimate aim. Another maintains that (SI) entails a false theory of personal identity. It entails that there is an enduring self through one's lifetime. What is true is rather that there is a succession of self-phases.

Parfit's second argument is the easier to refute. For (SI) does not in fact entail any theory of personal identity. It says that a person should act for her own good. What her good is depends on what the self is. Yet (SI) contains no criteria specifying what a person's good is, and it contains no commitment to a particular view of the self. So, none can either support or attack it. Only the applications of (SI) change when it is projected onto different theories of the self.

For instance, suppose that Parfit's account is true. Suppose I am successive overlapping replicas—S1, S2, S3, and S4—and that I am

⁷ Parfit gives the personal identity argument in *Reasons and Persons*, pp. 312-318. The achievement argument is on p. 133. He makes it part of the theory that a person has most reason to do whatever would best fulfil her present desires, but this inclusion is arbitrary. For what makes it reasonable to want accomplishments has nothing to do with their being present aims and what differentiates them from (direct) self-interest has nothing to do with time.

now in S2. S1–S4 are phases of the same person in the way succeeding generations are parts of the same nation. Then I have less reason to provide for the future than I once thought. This is a novel conclusion about what concerns a self-interested person should have. I used to think I was an identical self throughout my lifetime. Parfit gives me a choice between being the confederation S1-S4 and being S2. So, I think I am now really S2; this is why I care less about the others. "My life that matters" is the duration of S2. If the word 'self' is pre-empted to cover S1-S4, I am forced to a new terminology for what I formerly called a *self-interest theory*. It is a "me-first" theory, but it expresses the same principle of choice as before: I choose a kind of life because it is a good for me.

Parfit thinks that (SI) requires equal concern for all parts of future time and explains that this does not make sense if the unity of a person is like that of a nation. In effect, he builds into his version of (SI) a reference to an identical self. He has a reason for doing so. He uses this reference to separate his version of (SI) from the principle that what is reasonable is to fulfill one's present desires (the presentaim theory). This is a good enough way to make the distinction if there is an identical self. A true theory can be the basis for further distinctions. On the other hand, if there is no such self, the only conclusion is that the line between (SI) and the present-aim theory must be redrawn—perhaps so that the former is about the duration of a replica-stage and the latter about a present within that duration. The conclusion is not that (SI) is false.

Parfit thinks (SI) entails an account of personal identity, because he uses reference to an enduring self to distinguish (SI) from a present-aim theory. Yet it is not necessary to use that particular device nor to put any metaphysics into (SI). If I stay up late at night and read one metaphysical theory past my limit, I change my ideas about what my advantage is. I do not change the idea that I should act for my advantage. (SI) does not entail any account of personal identity.

I turn to Parfit's other argument. He notes that a person acts reasonably, but apparently not from self-interest, if she aims her life at an achievement. Bishop Butler⁸ thought that every action directed

⁸ Sermon XI, in *British Moralists*, L. A. Selby-Bigge, ed. (Indianapolis: Bobbs-Merrill, 1964): "The most intelligible way of speaking of it seems to be this: that lar affections towards external objects, and the actions done in consequence of to find rational motives other than self-interest. Many philosophers who hold that advantage is only one among several kinds of motives repeat Butler's mistake. They

to a state outside the agent had a non-self-interested aim. This is not quite right. For an action is self-interested if its aim is a good for the agent. So, she does not have to want an external object under the description 'a good for me' but instead under that description which makes it a good for her—for instance, 'money in the bank'. All the same, Butler and Parfit are right that people act and even live reasonably for purposes that do not seem to benefit them. The lives of saints and explorers, for instance, are a challenge to (SI). Reason can approve of lives centering on kindness or discovery.

To discuss this sort of example, I take up a distinction, for which I have coined terms, between *direct* and *reflexive* purposes. An action is directly self-interested if *its purpose* is a good (as a means or as an end) for the agent. It is reflexively self-interested if *his doing it* is a good for him, whether its purpose is or not. Some examples will

make these definitions intelligible.

Consider these three actions: (a) working to make a profit for a firm, if you have invested in the firm; (b) cheering the success of a baseball team, if you are a fan; and (c) dedicating yourself to a cause, if you are a revolutionary.

(a) is a directly self-interested action, since its purpose is a means to the success of the agent's investment. On the other hand, (b) is not directly self-interested. The team's victory does not return a profit to the fan. Matters are not as if he owns the hot dog concession, so that his business waxes or wanes with their fortunes. Rather, it is their success and not his own for which he cheers. Now, someone might object: "He identifies with the team. He has an emotional investment in them, and their fortunes are a means to his (emotional) profit." Which is true, but the fact that he identifies with them means only that he takes their success to heart. He feels toward it as if it were his own. There is in this no trace of direct self-interest.

Why do people take sides in professional sports? Cheering a team is not altruistic, selfless, magnanimous, public spirited, or dutiful. Presumably, the answer is: "For the excitement. Identifying with the team's fortunes stirs psychic waves that they cannot generate out of their private existences." Taking sides is then a reflexively self-inter-

note that we usually do not act from the intention, "This will advance my self-interest," and infer that we then do not act from self-interest. Yet we so act if we intend things under the aspects that make them goods for us.

There is also a saying, passed from writer to writer like a picture of a cute baby, that to be successful in pursuing our advantage, we would do well *never* to act from the intention, "This will advance my self-interest." The idea is preposterous, as though we had always to approach our ends sideways, without looking at them. All that makes sense is that self-interest requires us to aim at specific goods.

ested action. Doing it is a good for the agent, although what it aims at (the team's success) is exclusively someone else's good. (Notice that the distinction is not between means and ends—the team's good is not a means to the fan's good. Rather, cheering the team is his good.)

People often read fiction for the same kind of reason. Suppose someone is touched by the romanticism of Prince Andrew's life in War and Peace. This is directly an exercise of sympathy, although not motivated by altruism. It is reflexively, as an emotional gratification for the reader, a part of self-interest. Following professional sports and reading for pleasure are concerns for another's good undertaken from self-interest. It is not at all believable that they are instances of altruism, and yet sometimes their very content is concern-for-another. The distinction between direct and reflexive interest makes intelligible how this is possible.

Example (c), dedication to a cause, has about it an air of moral nobility which (b) lacks, but the two can have exactly the same relation to self-interest. Dedication is directly a selfless action. The revolutionary might do it for the same sort of reason a fan does—it gives a purpose to his life which he cannot find in his private concerns. Those are too insignificant, he thinks; a meaningful life requires a worthier cause. It is a reflexively self-interested action. This does not make it selfish, nor does it make it mean.

The point to which the distinction moves is that a selfless life is one whose direct aim is selfless. If it is also reasonable, its reflexive aim is self-interested. I am not saying that the agent's motive must consist entirely in the reflexive good, but I am saying that a selfless direct motive makes no sense by itself. The old reason comes forward again: without reflexive self-interest, either selflessness is directed to something that is not a good, and therefore marks a wasted life, or it subordinates the agent's good to that of others, and shows servility.

The objection to this must spring to everyone's mind. It is that the distinction is useless for supporting the belief that self-interest can be the purpose of a moral life. If something is done on account of its direct selfless aim, then it is not at all motivated by self-interest. No doubt, a person will obtain a satisfaction from doing it, but, as Bishop Butler said and as my definition allows, that an action is satisfying does not mean that it is self-interested. On the other hand, if it is done on account of its reflexive self-interested aim, then it is not motivated by moral obligation, as Kant would insist. So, the dilemma: either the motive is not self-interested (Butler) or it is not moral (Kant).

The answer to this is that a direct and a reflexive aim can have two kinds of relation. Sometimes the nature of the reflexive aim is inde-

Digitized by Arya Samaj Foundation Chennai and eGangotri

pendent of that of the direct one. For instance, revolutionaries are brought into cooperation, and the amiability of working together is a reflexive good. It does not depend on the end for which the group is working. Someone who wants only this amiability is not moved by the social good. Other times the nature of the reflexive aim derives from that of the direct one. The law has different direct purposes. A public defender intends to protect the rights of the accused; a corporate lawyer usually works to keep property in the hands of those who have it already. Both activities have many of the same reflexive qualities —the work is enjoyable as an exercise of knowledge, of mental skills, and of qualities of person. Still, keeping the rich rich is not much of a purpose. So, someone might reason: "Because this direct aim is not significant in itself, furthering it cannot be an accomplishment for me. A career devoted to it and one devoted to public defense have different reflexive qualities because their direct aims are different. One is not meaningful for me and one is." Now, to have a meaningful career is part of anyone's self-interest. Further, since what makes it meaningful to the public defender is the justice of its direct purpose, she deserves moral credit for pursuing it. There is no gap into which the Butler-Kant dilemma can fit.

I cannot leave my reply here. The quality of being an achievement for the agent has an unstable look. If it is an action's quality, it threatens to disappear into the direct aim. It would if 'x is an achievement' meant "x's direct purpose is significant." If it is not the direct aim, it might seem to migrate from the action and become part of the agent's satisfaction. So, it is unstable. It is in danger of slipping away from being an action's quality into being either its purpose or a feature of the agent.

Now, it is not identical with a direct aim. Something can have an important end without being important to a particular agent. Further, the revolutionary looked for a large social purpose so that his actions would be significant. Since the first is a reason for the second, they are not the same. An accomplishment is also not the same as an enjoyment. For something can be satisfying on other grounds—because it is pleasant. Also, that it is an achievement is for the agent the reason why it is satisfying, and, so again, they are not the same thing.

I do not want to begin a discussion of what makes something an achievement. That is a sober problem in real life and for an ethics that sees right action as part of a person's well-being. Here I have wanted only an example to make a formal point. A reflexive interest can be so tightly bound with a direct one that the Butler-Kant dilemma does not apply. This particular example also gives a twist to my argument: Parfit said that an achievement can be reasonable and

Digitized by Arya Samaj Foundation Chennai and eGangotri

selfless, but being an achievement is itself a quality that makes an action reflexively self-interested.

The distinction between a direct and a reflexive purpose is between the good to which an action is directed and the action as a good for the agent. It was drawn to defend (SI). It does so by supporting the position that, where a person's direct life aim is selfless. her reflexive one must be self-interested if she is reasonably motivated. There is an objection to using this distinction, not to defend (SI) in itself, but to defend the conjunction of (SI) and (D): if a direct selfless aim is a person's motive, she is not self-interested; if a reflexive interested one is, she is not dutiful. Reply: the reflexive one sometimes gets its nature from the direct one and, therefore, cannot be separated from it. The bottom line is that what Parfit gives as counterexamples to (SI)—achievements—are only apparent.

Parfit's argument based on self-identity failed because it reversed the conceptual ordering between "self" and "self-interested." The achievement argument failed because, although many accomplishments are directly non-self-interested, they must be reflexively selfinterested to make sense.

So, none of the positions, from Kant onward, contains a workable alternative to (SI). Are there others? There might be. I gave a reason, but not a conclusive one, to think that principle and sympathy are the only two classes of possibilities. Someone might also think that the issue turns on finding an unexplored variation within one of the classes. Now, the pressure to find an alternative is generated by this argument:

If (SI) is true, then, if anything is a motive for a moral life, it is self-interest.

There is a motive to live virtuously.9

⁹ Of course, this claim needs a defense. For the present, it gets only an introduction: suppose the extreme form of its denial—that an agent has two life plans, one of which is categorically right for her to follow (a virtuous life) but for which she has no incentive at all, and another which completely motivates her but for which there is no justification at all. On what grounds can she now decide how to live? She cannot regret selecting the completely motivated life, whatever harm she causes others, since she has no motive to choose otherwise. She can derive no satisfaction from choosing the other life, although she categorically ought to do so. Here obligation and satisfaction, and guilt and regret on the underside, do not overlap. Decision is simply unmotivated on the one hand and unjustified on the other. She has nothing to deliberate about. Her choice is absurd. Correspondingly, then, a choice of how to live is rational so far as it is both motivated and justified.

What would a senseless choice be like? I can imagine that some people see a decision between their religious beliefs and their sexual preferences almost absurdly. Their whole sense of duty moves toward religion, but without pulling along any of their desires. They find themselves drawn to a form of sexuality lying outside

It cannot be self-interest—(D), the claim that a virtuous life is a good for the person who leads it, is false.

So, (SI) is false.

There can be no conviction that (SI) is true if (D) is hopeless. It is because (D) is so in the eyes of post-Kantians that their efforts to solve the problem of moral motivation have been directed to alternatives to (SI). If (D) is not, and if my arguments so far have been correct, then our efforts should be shifted to defending (D).

V. THE DEFENSE OF (D)

A large argument for this shift is the fact that the attacks on (SI) have really supported (D). Kantian self-respect is a part of welfare. So, in arguing that it can motivate a moral life, he is arguing that an aspect of self-interest can—that is, for (D). Nagel shows that it is reasonable to be moved by the good of others, as morality requires us to be sometimes moved. Now, a happy life requires the same thing, since a selfish one is grubby even for the person in its center. So, Nagel has argued for an overlap of morality and advantage-again, for (D). Similarly for Parfit: he notes that people have reasonable desires for selfless achievements. They do, because an achievement is a kind of success, and a successful life is central to anyone's happiness. (J. L. Austin remarked somewhere that "success" is not a bad translation of the eudaemonia at which Aristotle thought everyone aimed.) Parfit shows that a satisfying life, like an exemplary one, can have a direct aim that is selfless—a lemma towards (D). The use of all these arguments is to explain how self-interest either includes or is compatable with something that at first appears to be non-self-interested. Their solid result is to support (D) rather than to establish an alternative to (SI).

anything permissible for them, so that they despise themselves even for feeling drawn. There is no point in deliberating about a choice seen in these terms. People can make no sense of it, unless either conversion warms them to their faith or enlightenment frees their conscience.

Plato said this in Laws, when he imagined a child asking its parents: Which would you rather, that I be happy or that I be moral? His implication is that the alternatives form a dilemma. Now, in real life, people often see themselves as presented with a choice between happiness and morality. They do not regard it as insoluble, because they do not see it as one in which all their motives line up on one side and all justifications on the other. Rather, they reach a decision precisely by reminding themselves of the motives they have to act morally—reputation and self-esteem are at stake—so that morality does not strike them as a wanton obligation. People remind themselves also that they cannot be unjustified in claiming a chance at happiness, that a modest self-respect entitles them to that much. They see the choice as between courses which are differently motivated and both partially justified. But, if they saw the choice as Plato imagined it, between what they ought to do (without any motive) and what moved them (without any justification), it would be impossible.

Most people will be surprised. For almost everyone feels that Kant and the others are arguing on the right side: the philosophical task is to show that self-interest is not so complete a rational motive as it seems, and this mirrors the real life task of getting people to be less self-interested, so that the voice of conscience will move them. The arguments are like medieval ones for the existence of God. Everyone is drawn already to their conclusion. Without this predisposition, we can see at once the arguments are strained in being used to attack (SI). Their natural bent is to support (D).

They also indicate the strategy for a successful defense. It must consist largely in an analysis of the concept of "self-interest." The idea is not that we should look for neglected facts about the benefits of a moral life. It is that we need conceptual clarity about what is advantageous. Interest is directed to actions and states of person as much as to external objects. Further, pursuing one's own does not require treating other people as means.

If it is directed to states of person, then self-interest is consistent with morality's requirement that action be principled. Stoicism develops this line of thought, and Kant draws his ideas about self-respect from it. (No doubt, Stoic tradition is too flamboyant, since it says that character is a person's only reasonable interest, but a milder idea is credible.) If a self-interested person can see another's good as an immediate rather than an instrumental concern, then self-interest is consistent with the moral demand that others be valued for their own sakes. The discussion of Nagel and Parfit goes some distance toward this. Marxist theory goes farther. It maintains that basic goods are public; so, self-interest is not exclusive. We would have a different concept of advantage if we were detached like the Stoics or revolutionary like the communists. A defense of (D) from their viewpoints presents an open vista. ¹⁰

A smaller part of the strategy consists in purifying the content of morality. Conventional and many rational moralities go against self-interest, but for different reasons. The former, along with its honest core requirement of human decency, also commands obedience to authority and deference to property. For instance, the rules about contracts are both honest, because keeping one's word is part of integrity, and fraudulent, because they are biased toward people with property. The prohibition of extramarital sex contains the same blend—it enforces responsibility for raising children, but it also treats sexuality as property and, like many rules governing the in-

¹⁰ I have written about these two topics in "Self-Interest and the Concept of Morality," *Noûs*, XXI (September 1987): 407–419.

stincts, is a device for maintaining obedience. Subservience to property and authority cannot be self-interested, and therefore not all of conventional morality can be, but there is no reason to think that rational ethics will make such demands. Their difficulty, rather, is that they often command altruism. They do so because they take it to be necessary in order to value other people for their own sakes. This thought returns us to the main line of defense, a re-examination of the concept of "self-interest."

I am not defending (D) now. I am giving reasons for thinking it is defensible. The first is that arguments against (SI) make sense only when pointed in that direction. The second is that Stoicism and Marxism are a viewpoint on which it is true. The relevance of discussing (D) at all is that the conjunction of (SI) and not-(D) entails that there is no moral motive. This entailment is not an argument against (SI) only if (D) is defensible. I have defended (SI). I gave the intuitive case for it in section III and showed in section IV the failure of efforts to find an alternative. (SI) appears to me, as it did to the ancient philosophers, to be the initial truth which must be accepted in discussions of moral motivation.

When he was deliberating about how to live, St. Augustine asked, "What does anything matter, if it does not have to do with happiness?" His question requires explanation, because he is not advising selfishness nor the reduction of other people to utilities, and even qualification, because other things can have some weight. All the same, the answer he expects is obviously right: only a happy life matters conclusively. If I had a clear view of it, I could have no motive to decline it, I could regret nothing by accepting it, I would have nothing about which to deliberate further.

STEPHEN SCOTT

Eastern Washington University

BOOK REVIEWS

Cogitations. JERROLD J. KATZ. New York: Oxford University Press, 1986. 206 p. Cloth \$24.95.*

Given the central role of the first person, subjective point of view in Descartes's philosophy, one would expect that the intense recent work in semantics on the so-called indexicals, like 'I', would have shed new light on Cartesian arguments like the cogito. Such expectations have not been entirely satisfied, however, due in part to the increasing narrowness and lack of epistemic motivation of much of the new semantics, and in part to the fact that Descartes has become something of a bête noire from whom philosophers—wary of dualism, the subjective point of view, and the threat of skepticism and solipsism—have been only too anxious to distance themselves. Jerrold Katz's new study of the cogito, Cogitations, in opposition to this philosophical climate, offers a refreshing new line of approach from the point of view of a distinguished philosopher of language. What is distinctive about Katz's approach is, to put it bluntly, that Descartes is taken seriously. Instead of bringing to bear the latest fashions in philosophy of language, epistemology, and philosophy of mind, almost all of which are distinctly anti-Cartesian in spirit, Katz, with an ear for ancient voices, tries to listen afresh to what Descartes has to say. Indeed, Katz's methodology may be his most important contribution, for instead of using current logistical and linguistic theory (which, appearances notwithstanding, was given to us by Gottlob Frege and W. V. O. Quine, not by God) to "correct" Descartes, he seeks the appropriate logico-linguistic framework in which to reconstruct Descartes's ideas and thereby enable one to recover the philosopher's insights and assess his limitations.

The principal goal of the book is to resolve what Katz calls the Cartesian Scholar's Dilemma. This is the problem that the inference 'I think, therefore, I am' is not a valid logical inference as it stands (being of the form: p, therefore, q), but that both obvious avenues of resolution are unsatisfactory. It is unsatisfactory to hold that as great a mathematician as Descartes simply committed an elementary logical mistake here, as well as to maintain that as clear-headed and lucid a thinker as Descartes simply omitted to fill in the missing step in the inference: 'whatever thinks, exists' (a step, moreover, which the philosopher always denied was required). Katz argues persuasively that

0022-362X/88/8509/0500\$00.60

© 1988 The Journal of Philosophy, Inc.

^{*} I would like to thank Professor Katz for discussing his book with me and for giving me the benefit of his comments on this review.

this simple but embarrassing dilemma has never been adequately resolved. His own way out is to recommend that we view the cogito as an example of analytic entailment as opposed to logical inference. He presents his own intensional semantic theory in which he distinguishes sharply between analytic entailments, where the sense of the predicate is "contained" in that of the subject, and synthetic ones, where the validity is due to a necessary connection between the relevant concepts, be they logical, mathematical, physical, or what have you. The former he attributes to the (sense) structure of language itself, the latter to the necessary links due to the concepts at hand, captured by the relevant theory of such concepts. On this account, then, Katz can explain why Descartes could claim that the cogito was not, after all, a suppressed syllogism, but rather something visible to a clear "intuition," since in thinking 'I think', on this

approach, I am already (tacitly) thinking 'I am'.

Katz's study divides into three main parts. In the first of these, he presents, in some detail, his neo-Fregean, neo-Chomskyan, "intensionalist" theory, in which surface, syntactic structure masks the deep, semantic structure of the language. In this part, he also defends his theory against the legendary assaults of Quine, who has argued that a principled analytic/synthetic distinction cannot be drawn. Basing his own reasoning on a post-Chomskyan linguistic framework, broader than the one that Quine employs, Katz marshals a powerful case. He also defends his own "concept-containment" notion of analyticity against the alternative account stemming from Kant and Frege, which connects analyticity with definition and logical inference. In the next section, he defends the application of his linguistic theory to the cogito. Here he argues that 'exists', pace Kant, Frege, et al., is a genuine linguistic predicate, though a "nonampliative" one (i.e., one that does not "extend" the concept of the subject). He then defends the cogito by arguing that the position of 'I' in 'I think' is "referential," i.e., that the 'I' purports to refer, so that 'I think' (like 'I drink coffee') presupposes or implies that the 'I' succeeds in referring, and, therefore, that its referent exists. In the final section of the book, Katz defends the referentiality of the Cartesian 'I' from (Wittgensteinian) attacks by Elizabeth Anscombe and Peter Geach. He then assesses the significance of the cogito in establishing the existence of a thinker (at best, momentary, he believes) and in withstanding Cartesian doubt. Analytic entailments like the cogito, he urges, are immune from the deception-game of the First Meditation, since one cannot even contemplate what it would be for them to fail (as this would be to violate the linguistic rule of supposition itself).

As can be seen from this brief summary, Katz's book is bold and ambitious, not only in offering a new account of the cogito, but also in presenting his intensionalist semantics and taking on powerful contrary approaches by Kant, Frege, Wittgenstein, and Quine in the process. It is impossible in a short review to assess so ambitious a project adequately. Although I am sympathetic with Katz's approach to the cogito, as with various aspects of his approach to semantics, I cannot give either the attention it deserves. Most impressive, I found, were his account of the cogito as an analytic entailment, his defense of analyticity against Quine's attack, and his argument, against Anscombe and Geach, for the referentiality of the Cartesian 'I'. His response to Quine can stand alongside other recent and forceful criticisms of Quine's philosophy by Hao Wang1 and Barry Stroud.2 Although I cannot discuss the semantic theory that underlies Katz's new approach in detail, there is room to mention a few specific points that concern me.

Katz focuses his theory of "senses" on what might be called the "horizontal level" of the connections among senses themselves—on synonymy, antonymy, homonymy, heteronymy, entailment, containment, etc. He provides strong support for his thesis that we have intuitions about such sense-sense connections, and that they are a priori and concern a Platonic, independent realm of entities, not our human psychological capacities. The full case for this is made out in his Language and Other Abstract Objects,3 where he also develops a suggestive analogy with mathematics. (An excellent shorter presentation, and impressive defense against counterattacks, is found in his "An Outline of Platonist Grammar." Detailed differences between his theory and Frege's are outlined in "Why Intensionalists Ought Not Be Fregeans." One salient difference, repeated in Cogitations, is that Katz rejects Frege's doctrine that sense gives the "mode of presentation" of the referent. He rejects the idea that sense determines reference. For Katz, sense determines only "typereference," reference in the language, not the "token-reference" of specific uses of referring terms in particular contexts. He hopes, using this distinction, to avoid the famous "counter-examples" to

^{1 &}quot;Two Commandments of Analytic Empiricism," this JOURNAL, LXXXII, 9 (September 1985): 449-462.

² "Naturalized Epistemology," in his The Significance of Philosophical Scepticism (New York: Oxford, 1984), ch. 6.

³ Totowa, NJ: Rowman & Littlefield, 1981.

⁴ In Katz, ed., The Philosophy of Linguistics (New York: Oxford, 1985), pp. 172 - 203

⁵ In Ernest LePore, ed., Truth and Interpretation: Perspectives in the Philosophy of Donald Davidson (New York: Blackwell, 1986).

sense theories of reference due to Keith Donnellan, Hilary Putnam, et al., in which one succeeds in referring to a particular object via a term whose sense, classically, does not apply to the purported referent. Thus, as in Donnellan's "Reference and Definite Descriptions," one is supposed to be able to succeed in referring to someone, as 'the man drinking a martini', even though the sense of the description fails to uniquely determine him. The "referential" use of a term, for Donnellan, can enable us to refer even though the "attributive" (sense-based) use would fail.

Now, what bothers me about Katz's type-theoretic account of sense is that it is not clear (a) how one is supposed to fill in what else is needed, to achieve actual, "token-reference," and (b) whether the missing ingredient will conform to the requirements of a Cartesian theory. In regard to (a), note first that Donnellan never really gave an account of the mechanism of the referential use of descriptions.7 In later papers he did offer an account of sorts of the mechanism of reference of proper names, a non-Fregean, causal/historical account. In "Why Intensionalists Ought Not Be Fregeans," Katz employs aspects of such a theory of the reference of proper names, but two problems arise: (1) On Katz's account, it is the speaker's knowledge of the causal/historical ancestry of a proper name which helps determine the referent, a Cartesian requirement precisely rejected by Donnellan, as well as by Saul Kripke, Putnam, et al., of the "new theory of reference"; and (2) this account of proper names will not help to explain the mechanism of reference appropriate for indexicals, like 'I', whose reference is determined not by historical factors but, rather, by the contextual features of the actual token speech act in question. We are thus in the dark both as to what the "extrasense" conditions are that determine the reference of a particular use of 'I', for Katz, as well as to whether these factors will turn out, in the end, to be in the Cartesian spirit of Katz's theory. This, it seems to me, is no small issue. Katz's "horizontal" account of sense is well-designed to help him explain the inferential move from 'cogito' to 'sum'. But the Cartesian philosophy, as he is well aware, also seeks an Archimedean point, the existence of the self, on which to build up one's world. The 'I', for Descartes, is a tangent point at which thought touches reality, and here we are in need of the missing "vertical" part of Katz's semantic theory, the part where a bridge is made from sense to reference. Without a precise account of how

⁶ The Philosophical Review, LXXV (July 1966): 281–304.

⁷ For a sober and intelligent assessment of Donnellan's essay, see Mark Sainsbury, Bertrand Russell (Boston: Routledge & Kegan Paul, 1985), pp. 122–126.

my specific use of 'I' gets me to myself, the full force of the cogito in Descartes's "subjective" philosophy cannot be understood and evaluated.8

A solution to this question is crucial to coming to grips with the Cartesian project of establishing the existence of an "external." physical world, including, and especially, one's own body. Katz, as we have seen, discusses Geach and Anscombe on 'I'; but I would also like to see him discuss Peter Strawson's Kantian-Wittgensteinian attack on this aspect of Descartes's philosophy. For Strawson, too, believes that sense alone does not determine reference, but that an irretrievably indexical, contextual aspect is also required. However, Strawson, unlike Katz, is led to a very un-Cartesian account of the self and 'I'. For him, I can succeed in referring with 'this body', though its referent be epistemically indistinguishable from a body in a "twin" part of the universe, but 'I' cannot, for Strawson, be used similarly to refer successfully to "this mind" (as opposed to its "epistemic twin"). One would like to see Katz's response to this line of thinking, as well as to Strawson's combination (in Individuals), by brute force, of "Cartesian" with bodily, "external" aspects of personhood, into the primitive yet complex concept of the person. To my mind, we have, to date, no adequate synthesis of these different aspects of personhood, nor, indeed, a full account of the relationship of the semantics and epistemology of reference to the Cartesian project. In fact, things have gotten into such a mess that we find, for example, Kripke, in one and the same work (viz., Naming and Necessity10), stoking the fires for the anti-Cartesian semantics of the "new theory of reference" as well as, in the final sections, mounting one of the few defenses around, from a modal point of view, of the Cartesian theory of the essential nonphysicality of mind. One would have liked to see as sustained and intense a study of the Cartesian 'I' as Katz's touch on such issues.11

Finally, I would like to comment on two other aspects of Katz's account. The first is that he must square his intensionalist, Platonist

⁸ Such an account is hard to come by. I have argued elsewhere that even the well-known Kaplan/Perry theory of indexicals itself stands in need of a "bridgetheory" to show how, in the end, we achieve "cognitive access" to our referents. See "Frege, Perry, and Demonstratives," Canadian Journal of Philosophy, XII, 4 (December 1982): 725-752.

The Bounds of Sense: An Essay on Kant's Critique of Pure Reason (London: Methuen, 1966); and Individuals: An Essay in Descriptive Metaphysics (Garden City, NY: Doubleday, 1963).

Cambridge: Harvard, 1980.

For an extended discussion of the relationship of the new semantics of Kripke, Kaplan, et al., to classical Cartesianism, see my "The Path Back to Frege," Proceedings of the Aristotelian Society, New Series, LXXXVII (1986): 169-210.

account of language in the rational reconstruction of the validity and necessity of the cogito, with Descartes's view that necessity rests, in the end, not in any abstract structures but in the free choice of God. Second, even if one grants Katz that there is a special ground for the immunity from doubt of analytical entailments such as (on his view) the cogito, based on the very meaning of the terms in the supposition, is it so clear that going beyond this to the question of the content of a supposition necessarily weakens our epistemic position, even where the content is, say, a primitive arithmetic one, about which our intuitions would appear to be conclusive? This raises, I think, an interesting epistemological issue, in addition to the one Katz has explicitly concerned himself with. He has presented one way to achieve demon-proofness, the way of the cogito; may there not be others? To my mind, whatever one makes of the slippery notion of certainty, there is surely nothing, including the cogito, that is more certain than are simple arithmetic truths like '2 + 3 = 5' (whatever may be the source of one's certainty about them). If someone were to insist that an evil demon could have made me believe any (nonanalytic) nonsense whatever, including arithmetic nonsense, I would respond by introducing the following distinction: (1) An evil demon could have made me believe nonsense arithmetic. This I agree to. (2) Concerning my current beliefs (including '2 + 3 = 5'), an evil demon could have made me have them (though all turn out to be nonsense). This I reject. I reject (2), because (2) implies (3): possibly, $2 + 3 \neq 5$, which, if one is to speak of certainty at all, is quite certainly false. Of course, this argument itself rests on a logico-arithmetic theory, L, which, if Katz is right, it at least makes sense to reject (unlike the cogito), and thus which one can meaningfully suppose the demon to have misleadingly foisted on us. But I would insist that to entertain such a doubt is, again, to suppose that: (4) possibly, L is not sound; and this I would certainly reject. As William Tait12 says about finitary reasoning, "it is indubitable in a Cartesian sense that there is no preferred or even equally preferable ground on which to stand and criticize it" (ibid., p. 525).

Cogitations, then, is not the final word on the cogito, but it is a much needed re-evaluation of the Cartesian philosophy. I hope that its provocative insights, as well as its unanswered questions, prove an impetus to more scholarship on this high level of sustained seriousness.

PALLE YOURGRAU

Barnard College, Columbia University

^{12 &}quot;Finitism," this JOURNAL, LXXVII, 9 (September 1981): 524–546.

NOTES AND NEWS

The editors report with sorrow the death of Emerson Buchanan, Professor Emeritus at Fairleigh Dickinson University and former managing editor of this JOURNAL, on May 29 in Pella, Iowa. At the time of his death he was 78 years old. In 1952, Professor Buchanan succeeded Iris Wilder Dean as managing editor of the JOURNAL and, in the ten years that followed, established the meticulous editorial standards that the JOURNAL has tried to maintain.

From 1927 to 1931 he was a Pulitzer Scholar at Columbia College; he received his AB in 1931 and his PhD in 1959, and, in 1960, the Woodbridge Prize for Distinguished Research. During his tenure as managing editor, he was also head of the Philosophy Library at Columbia. In 1962, he went to Fairleigh Dickinson University, where he taught philosophy until his retirement in 1976.

The editors are grateful to Professor Buchanan for his years of service at the JOURNAL and extend their sympathy to his wife, Valerie Buchanan, and to his niece, Professor Phyllis Cassidy of Smith College, formerly business manager of the JOURNAL.

The Philosophy Department of the University of Missouri/Columbia will host two conferences on historical and contemporary perspectives on the representational nature of thought. The first conference, to be held October 13–14, will be devoted to the views of Descartes and Kant. Participants will include Henry Allison (San Diego), Karl Ameriks (Notre Dame), Richard Aquila (Tennessee), Janet Broughton (Berkeley), Calvin Normore (Toronto), Hoke Robinson (Memphis State), James Van Cleve (Brown), and Margaret Wilson (Princeton). The second conference, to be held November 17–18, will be devoted to contemporary views. Participants will include Fred Dretske (Wisconsin), Hartry Field (Southern California), Jerry Fodor (CUNY), and Robert Stalnaker (Cornell). Further information may be obtained by writing Peter Markie, Phil. Dept., 438 GCB, UMC, Columbia, MO 65211.

The Social Philosophy and Policy Center would like to announce its third and final conference on the topic "Economic Rights and the Constitution," to be held at Bowling Green State University on October 20–22. Participants will include Frank I. Michelman (Harvard), James Buchanan (George Mason), R. Shep Melnick (Brandeis), Stephen Macedo (Harvard), Leo Troy (Rutgers), William Riker (Rochester), Thomas Haggard (South Carolina), Mark Tushnet (Georgetown), and Lino Graglia (Texas). Information may be obtained by writing Kory Tilgner, Social Philosophy and Policy Center, Bowling Green State Univ., Bowling Green, OH 43403.

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 10, OCTOBER 1988

Gurukut Kangri Vishwar, tvaray HARBWAR

ON THE STABILITY OF THE LABORATORY SCIENCES*131

S our knowledge of nature unstable, constantly yielding to refutation or revolution? You might think so, to judge by recent philosophy. In this summary I propose a framework in which to understand the manifest fact that science since the seventeenth century has, by and large, been cumulative. The framework does not question the chief insights of Karl Popper or Thomas Kuhn, but places them in a larger perspective.

My theme is the stability of "mature" laboratory sciences. It concerns only those sciences in which we investigate nature by the use of apparatus in controlled environments, and are able to create phenomena that never or at best seldom occur in a pure state before people have physically excluded all "irrelevant" factors. By laboratory science I do not mean just the experimental side of a science. My topic is the stabilizing relationship between theory and experiment.

Many mature sciences are pedagogically stable. We learn geometrical optics when young, the wave theory as teenagers, Maxwell's equations on entering college, some theory of the photon in senior classes, and quantum field theory in graduate school. Newton's rays of light particles are omitted, as are many other byways. Each of these stages is taught as if it were true. The sophisticated teacher may add, "but not really true." Each earlier stage is at best approximately true.

Some scientific realists embrace this (and an implied stability), holding that science converges on the truth. The implication is that earlier stages are approximately true. Paul Feyerabend rejected the

 $^{0022\text{-}362}\text{X}/88/8510/0507\00.80

© 1988 The Journal of Philosophy, Inc.

^{*}To be presented in an APA symposium on The Philosophical Significance of Experimentation, December 28, 1988. Patrick Heelan will be co-symposiast, and Peter Galison will comment; see this JOURNAL, this issue, 515–524 and 525–527, respectively, for their contributions.

Digitized by Arya Samaj Foundation Chennai and eGangotri approximation idea on the grounds that successive stages are incommensurable. Before Nancy Cartwright, few on either side examined actual practices of approximation. One can argue that her multiplicity of possible approximations both toward and away from the truth suffices to show that all such considerations about approximation are jejune.

Here is a telling example due to S. S. Schweber. In 1981, workers at the University of Washington devised the Penning trap, which contains a single electron in a definite space. Everything they did was planned according to and can be explained by the prerelativistic (pre–Dirac) theory of the electron. It is not clear that it can be done otherwise. For those purposes, that old account of the electron is better than any other; it is the account which is true to the facts: true to the experiment and its applications.

One suggestive idea about how stability arises relies on the observation that theories and laboratory equipment evolve in such a way that they match each other and are mutually self-vindicating. Such symbiosis is a contingent fact about people, our scientific organizations, and nature. Another contingency is that new types of data can be produced, thought of as resulting from instruments that probe more finely into microstructure, data which cannot be accommodated by established theory. This creates space for a mutual maturing of new theory and experiment, but does not necessarily dislodge an established mature theory, which remains true of the data available in its domain. 'Data', 'theory', 'experiment', 'equipment'—these are familiar words, but, to expound this notion of stability, we require a finer taxonomy.

ELEMENTS OF LABORATORY EXPERIMENT

Thanks to a large number of recent studies by philosophers, historians, and ethnographers of experimental science, we have much richer sources of material about the laboratory than was available a decade ago. The welter of colorful examples makes it hard to produce any tidy formal characterization of experiment. Hence, our powers of generalization are limited. I shall try to return some degree of abstraction to the philosophy of science by listing some familiar elements in laboratory experimentation.

(1) There is a question or questions about some subject matter. (A question answered at the end of the experiment may not even have been posed at its commencement; conversely, what prompted the work may disappear not only from the official write-up, but even from living memory). Questions range from those rare ones emphasized by philosophers ("Which of these two competing theories is false?") to the commonplace ("What is the value of this quantity?" or

Digitized by Arya Samaj Foundation Chennai and eGangotri

"Does treating X with Y make a difference?"). When a question is about a theory, I shall speak of the theory in question.

- (2) Established or working theories or background bodies of knowledge or assumptions about the subject matter. These are of at least three kinds.
- (2a) Background knowledge and expectations which are not systematized and which play little part in writing up an experiment.
- (2b) Theory of a general and typically high-level sort about the subject matter, and which, by itself, may have no experimental consequences. Call this systematic theory.
- (2c) What in physics is commonly called phenomenology, what R. B. Braithwaite called Campbellian hypotheses, and what many others call bridge principles. I shall speak of topical hypotheses. Hypothesis is used in the old-fashioned sense of something more readily revised than theory. It is overly propositional. I intend it to cover whole sets of approximating procedures in the sense of Cartwright, and what Kuhn called the "articulation" of theory in order to create a potential mesh with experience. That still ignores a more tacit dimension, the skills used by the "phenomenologist" to create that articulation in practice. Topical is meant to connote both the usual senses of 'current affairs' or 'local', and also to recall the medical sense of a topical ointment applied to the surface of the skin, i.e., not deep. Here are two extreme examples from physics: in the case of a measurement of local gravitational acceleration (3 c below), the systematic theory is Galilean mechanics, and there is today no self-conscious "phenomenology," no formulated bridge principles or topical hypotheses. In the case of superstring theory, a potential grand unified theory of many dimensions, topical hypotheses connect this structure with something that happens in our three or four dimensional world.
- (3) There is the materiel of the experiment. Commonly this breaks down into three parts, each associated with a set of instruments.
 - (3a) There is a target that is prepared by some devices.
- (3b) There is apparatus that is used to interfere with the target in some way.
- (3c) There is a detector that determines some effects of the interference.

I arbitrarily restrict the word 'apparatus' to (b), and will use instrument generically for (a-c). In elementary chemistry one mixes two substances to observe their interaction. They are the target. I would call litmus paper a detector, but not human observation, which I put in (5) below. The apparatus of Atwood's machine for Digitized by Arya Samaj Foundation Chennai and eGangotri determining local gravitational acceleration is a tuning fork with a brush on one prong. It is dropped so that the brush sweeps out a curve on the detector, a plate of glass with whitewash on it. There is no target materiel. Or is it the gravitational field?

- (4) There are theories or at least background lore about the materiel. They help us design instruments, calculate how they will work, debug them, and run them. The phenomenology of each instrument may differ. Seldom (never?) is the phenomenological theory of an instrument the same as the theory in question (1) or the systematic theory (2b). It may overlap with the topical hypotheses (2c). For example, the theory of the tuning fork has nothing much to do with the theory of gravitational acceleration. Nowadays in big science, people who design detectors and people who prepare targets commonly have very different expertise.
- (5) Data generators. In the past these were usually people like George Atwood measuring the length of the successive inflection points of the curve swished out by the brush at the end of the tuning fork. Now we have printouts of automatic readings. A camera that takes micrographs from an electron microscope once was a detector, now it is a data generator.
- (6) Data: the physical, material records produced by a data generator.
- (7) Data processing: a catch-all name for distinct activities that have appeared at different stages in the history of science.
- (7a) Data assessment: e.g., the calculation of a probable error, using a formal procedure that is theory—neutral. There are also estimates of systematic error based on theories of the detector, apparatus, and target, and on deductions from topical hypotheses.

(7b) Data reduction: indigestible or unintelligible information is transformed by statistical techniques into manageable quantities or displays.

- (7c) Data analysis, well-described by Peter Galison for high energy physics: the "events" under study-preserved, e.g., as tracks on photographs—were once selected and given preliminary analysis by semi-skilled labor. With the advent of very fast detectors, a tape of events had to be analyzed by computer. The technicians were trained, and later the programs were written, in the light of both topical hypotheses and theories of the instruments. Computer simulation of missing bits of data, and image enhancement, provide further examples of data processing.
- (8) Interpretation of the (processed) data: in the simplest cases this is a single stage: a series of trials on Atwood's machine gives us a set of pairs of distances (between swishes) and times; we then com-

Digitized by Arya Samaj Foundation Chennai and eGangotri pute the gravitational acceleration g according to the Galilean formula $g=2s/t^2$. We need no topical hypotheses. Unfortunately, reasoning like that used for Atwood's machine is the model for too much philosophical discussion of theory and experiment. EXTENDING PIERRE DUHEM'S THESIS

How is an "acceptable" experimental result obtained? Duhem observed that, if an experiment or observation is persistently inconsistent with a systematic theory, you need not abandon the theory, for in principle you can revise the theory of the instrument. In his example, revise astronomy or revise the theory of the telescope. Now, if you did the latter, you would probably rebuild your telescope, creating a substantially different instrument! Andy Pickering importantly advances Duhem's idea by adding the materiel to the items that can be modified.

He regards the systematic theory, the instruments, and theories of the instruments as three plastic resources which the investigator brings into a mutual adjustment. His example has two competing theories in question: free charges come either in units of e, the charge of the electron, or else 1/3e (free quarks). In the background (2a) these are the only two possibilities. The materiel is an upgraded Millikan oil—drop device. The initial results of the experiment described seemed to show pretty much a continuum of minimum electric charges. The experimenter modified both his theory of the apparatus and the apparatus (he revised the phenomenological theory of condenser plates, and repositioned the plates). The ensuing data were interpreted as refuting the theory of free quarks.

Duhem emphasized my elements (2) and (4). Pickering restored us to experimental practice by adding (3). Robert Ackermann attends to yet other elements, adding not only the instruments (3) but also the data (6) and interpretation (8). Like Duhem and unlike Pickering, he has a passive attitude to instruments, treating them as if they were "off—the—shelf" devices in the way that a navigator would use a chronometer, or a cell biologist a nuclear magnetic resonance spectrometer. What were once experimental and thoroughly "plastic" instruments become reliable technology.

In Ackermann's account, instruments produce data that are literally "given." The data are not theory-laden; they are material artifacts, photographs, or inscriptions, the productions of instruments. Theory enters when they are interpreted. Science is a dialectical affair of fitting data and theory together. Data that at one time are just "noises" may later be interpreted by a new theory. Thus, after the theory of pulsars was in place, older astrophysical records were shown to be rich in evidence of pulsars. Those records were not

Digitized by Arya Samaj Foundation Chennai and eGangotri theory–laden, but their interpretation is, on Ackermann's view, a

matter of theory.

Thus, Duhem, Ackermann, and Pickering point to different kinds of interplay among some of the elements (1-8). In fact, all eight are plastic resources. We can (1) change questions, or more commonly modify or sharpen them in midexperiment. Data (6) can be abandoned or selected without fraud. Data processing (7) is almost embarrassingly plastic. Those familiar with the logic of statistical inference will be well aware of the hazards of data reduction (7b). Data analysis (7c) is plastic in an entirely different way. Its computer programs are highly susceptible to modifications in the theory of target, apparatus, or detector-not to mention material changes made in the way that those systems operate. One aim of developing a taxonomy such as (1-8) is to describe a complex pattern of adjustments which concludes with stable science resistant to revision.

MATURITY AND STABILITY

Here is a very liberal adaptation of Ackermann's picture of maturing science. A collection of kinds of instruments evolves, hand in hand with theories that interpret the data that it produces. Ackermann calls a collection producing data that comes to fall under a systematic theory an instrumentarium. As a matter of brute contingent fact, instrumentaria and systematic theories mature, and data uninterpretable by theories are not generated. There is no drive for revision of the theory, because it has acquired a stable data domain. What we later see as limitations of a theory are not even perceived as data. Thus, geometrical optics takes no cognizance of the fact that all shadows have blurred edges. The fine structure of shadows requires an instrumentarium quite different from lenses and mirrors, together with new systematic theory and topical hypotheses to interpret it.

But is not at most one theory true, the old mature one, or the aspiring new one that takes account of the additional data domain? The metaphysical doctrine of the unity of science demands that. But think instead of the theories being different representations of several aspects of "the same reality." Niels Bohr tried to mitigate the shock of such a way of thinking by invoking complementarity, but we should reach further than that. New sense is given to the idea of incommensurability: these theories are incommensurable in Kuhn's intended sense of "no common measure." The measure of the mature theory is its data domain, which it fits within tolerances of error; the new theory tries to measure up to a new data domain.

Philosophers from Susan Stebbing have mocked A. S. Eddington's remark that he had two tables before him. Well, there is only one

Digitized by Arya Samaj Foundation Chennai and eGangotri

table. But we should think out a semantics of representation, in which many commonsense beliefs about the table are literally true, as were many of the things said by 1920s physics, and as are those held by quantum field theory. There is one table, and many incommensurable truths about it.

So much I take to be a liberal extension of Ackermann's suggestions. Two things are importantly missing. One is the long haul of getting to a mature theory, which I believe is best described by an equally liberal extension of Pickering's account of apparatus and apparatus theory as a plastic resource. The other is the role of those elements, in my taxonomy, which Ackermann does not mention. The chief practical indeterminacy of science lies not in the possible things one can do with the world or in the possible systematic theories one may entertain, but in the manifold of models and approximations which constitute topical hypotheses. Topical hypotheses are underdetermined. A theory does not say how it will be articulated to mesh with the world, nor does processed data say how they will be interpreted by topical hypotheses. There are lots of ways to do it, and every phenomenologist has a battery of such techniques. Kuhn has importantly emphasized that learning a science is not learning the systematic theory but learning how to do the problems at the end of the chapter. A casual inspection of many textbooks will show that, aside from certain mathematical tricks of calculation, these problems are typically training in how to use what I have been calling topical hypotheses. A mature science achieves a canonical set of approximations, the glue that holds it together, and which enable us to say that the theory is true of its data domain. The pain in hardworking science is the construction of new topical hypotheses. That is the "puzzle" to which so much "normal science" is addressed.

THESES AND QUERIES

(1) Scientific realism. All that I have said is consistent with scientific realism about entities. My description of mature and successor normal science strongly resembles Duhem's antirealism about theories. I attribute to him the conception that nature, and even "mechanics" or "optics," is too complex to admit of a single unified description. One can at best aim at characterizing an aspect of parts of nature, and this is achieved by complementary mature theories that need not be commensurable. But 'aspect' is no longer merely a metaphor, for it is to be explicated in terms of the structure of instruments, processed data, and topical hypotheses.

(2) Truth. Do we need a new semantics for real science, one that is based on the locution, 'true to the facts'—not the facts about some

Digitized by Arya Samaj Foundation Chennai and eGangotri metaphysical world, but the facts about the phenomena created by

experimentalists?

- (3) Kuhnian revolutions. Not all scientific revolutions are Kuhnian, witness "the" scientific revolution of the seventeenth century. It has been argued that there was a scientific revolution in geology leading up to plate techtonics, but lacking the stage of crisis which precedes paradigm shift. Conjecture: revolutions with Kuhnian structure are of two sorts. One occurs when a paradigm is imposed on a preparadigmatic field. The other sort occurs in mature sciences precisely when a new instrumentarium generates data outside an established data domain. That fits the "function" that Kuhn finds for measurement in the physical sciences, and his own study of black body radiation.
- (4) Laboratory technology and the unity of science. When an instrument becomes an off-the-shelf device for one branch of science. it can often be incorporated, after painful adjustments, into another. X-ray diffraction, designed for crystallography, engenders molecular biology. The instrument theory of one science becomes built into the practice of another. Insofar as topical and instrument theories interact, there is a resultant unification of data domains, and hence an apparent unification or at least congruity among sciences.

IAN HACKING

University of Toronto

Digitized by Arya Samaj Foundation Chennai and eGangotri EXPERIMENT AND THEORY: CONSTITUTION AND REALITY*

OW in a constitutional analysis do theories relate to experimental phenomena, and how do both relate to nature or reality: this is the theme of my paper. Examples are taken chiefly from physics, e.g., Robert Milliken's oil-drop experiment and recent experiments in fundamental particles physics.¹

The activities of scientific research all contribute to one or other of two distinct research strategies, experimental and theoretical. In this paper, we ask two questions: (1) How are the objects-for-knowing generated by experimental strategies related to the objects-for-knowing generated by theoretical strategies? (2) Which counts for reality?

The method I employ is *constitutional analysis*.² This supposes that objects—experimental or theoretical—are prepared by the noetic activity of subjects in anticipation of being presented as objects-for-knowing. Experimental and theoretical noetic activity are as different as looking at the sun and looking at the sunbeam—one can only do one at a time, for, though related, they are different noetic activities.

The function of theory is to explain via a mathematical model how and why the practical procedures work for the experimenter, whereas the function of experiment is to authenticate the presence of the phenomenon in the practical procedures. I take 'theory' to mean the mathematical model closest to experimental praxis (Hacking, 216–218; Galison, 249–254).

* To be presented in an APA symposium on The Philosophical Significance of Experimentation, December 28, 1988. Ian Hacking will be co-symposiast, and Peter Galison will comment; see this JOURNAL, this issue, 507–514 and 525–527, respec-

tively, for their contributions.

I i am deeply indebted to many important studies by Harry Collins, G. Holton, B. Latour, Andy Pickering, Trevor Pinch, Steven Shapin, and others, of the historical, sociological, and psychological relationships between theory and experiment. Two deserve special mention: Peter Galison, How Experiments End (Chicago: University Press, 1987); and Robert Crease, The Second Creation (New York: Macmillan, 1986). In the philosophical literature, I refer especially to Ian Hacking, Representing and Intervening (New York: Cambridge, 1983); and Robert Ackermann, Data, Instruments, and Theory (Princeton: University Press, 1985); and my Space-Perception and the Philosophy of Science (Berkeley: California UP, 1983); but important work has also been done by others. [References in the text to these works are given in the form: (author, page numbers).]

² Semantic analysis studies meanings, whereas constitution analysis studies the origins of meanings. For an account of constitution analysis, see, for instance, Robert Sokolowski, *The Formation of Husserl's Concept of Constitution* (The

Hague: Nijhoff, 1964).

0022-362X/88/8510/0515**\$**01.00

© 1988 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri Before plunging into constitutional analysis, let me present my conclusions in a preliminary way under the form of an analogy: a theory is related to a phenomenon as a musical score is related to a musical performance. Theory alone can no more witness to the authentic presence of a phenomenon than can the score alone of a piece of music witness to the authentic presence of a musical performance: there are hosts of other relevant factors in each case. The phenomenon and the musical performance are each the realization of theoretical models or schemata that depend on social, practical, technological, hermeneutical, and artistic judgments that are local,

contextual, and immersed in cultural history. HUSSERL'S TRANSFORMATION THEORY OF THE PHENOMENON The roots of constitution analysis are in Kant, Hegel, and Husserl. Husserl was trained in mathematics and taught at Göttingen (1901-1916) during its "Golden Years." He was a colleague and friend of the great geometers, logicians, and axiomatizers who set the agenda for twentieth-century physics.3

Husserl applied Felix Klein's transformation theory of projective geometry to the analysis of perceptual objects. Just as a geometrical object lies somewhere between the subjectivity of the coordinate system and the objectivity of its representation in that system, so the perceptual object lies somewhere between the subjectivity of an individual viewing and the objectivity of the view that the individual gets. The geometrical object, for Klein, is the symmetry or invariant preserved under permissible transformations of the representation and of the coordinate system; the perceptual object, for Husserl, is the symmetry or invariant preserved under permissible transformations of the viewing system and of the viewers. The key to the analysis is, of course, to identify correctly the transformation group (or symmetry group) of the phenomenon, i.e., the group of transformations that permute the profiles among themselves while preserving the stability of the phenomenon.4

Whereas most theoretical physicists have directed their research toward simplicity defined in terms of the symmetries of a model, most experimental physicists have directed their research toward a comparable experimental symmetry, the stable synthesis of a scien-

⁸ Husserl was himself an axiomatizer—in anticipation of Patrick Suppes? Well, yes and no. See my "Husserl's Later Philosophy of Science," Philosophy of Science, LIV, 3 (1987): 368-390, for a commentary on Husserl's philosophy of science and a critique of it.

⁴ The importance of symmetries or invariance under transformation groups is borne out by high-energy physics. As Steven Weinberg says, "The Universe is an enormous direct product of representations of symmetry groups. It's hard to say it any more strongly than that" (Crease, 187).

Digitized by Arya Samaj Foundation Chennai and eGangotri tific phenomenon. Although the centrality of the model is stressed by all who take their inspiration from Göttingen, the importance of a robust and stable phenomenon is, in the long run, more important. Both Ian Hacking and Peter Galison make a move in this direction (Hacking, 222, 229; Galison, 259–260).

The notion of view, profile, or perspective, however, involves a relation between perceiver and perceived. Every transformation affecting the object is paired with a correlative inverse transformation affecting the subject which would have the same phenomenological outcome. This second group of transformations, initiated by the agency of the subject, has the structure of the inverse group, but this has the same structure as the group itself, since every inverse transformation is also a member of the group.⁵

The correlativity of the objective and subjective transformation groups is in turn the key to understanding the constitution of the phenomenon. The objective transformation group is then the law of the objective constitution of the phenomenon. Husserl called it the *noema* of the phenomenon (cf. Heelan, 6–8). The subjective transformation group is the acquired ability of the prepared experimenter to explore at will the noema of an object, and this is (what Husserl calls) *noesis*. Noesis is the law of subjective constitution of the object. One is led in this way into the analysis of perception as an active process and as a species of experimental performance. Noesis and noema then are correlative and (since a group and its inverse are the same) both are representations of the same basic transformation group.

Constitutional analysis attempts (1) to define this basic transformation group, and (2) to describe its noematic and noetic implementations. These noetic implementations always employ the human body and its sensory systems; they may, in addition, involve technologies such as instruments and the laboratory traditions that use them.

SCIENTIFIC PHENOMENA AS PERCEPTUAL OBJECTS
Can experimental phenomena be or become perceptual objects in the Husserlian sense? The answer depends on whether an experi-

⁵ In physics, the first group is called the active group, and the second—for constitutional analysis more interesting—group is called the passive group. Physicists call it the "passive" group because it seems to them that the passive group leads merely to redescriptions of the phenomenon. Husserl would have pointed out that observers are not passive spectators, that the phenomenon itself subsists between perceivers and profiles, and that, as a consequence, a constitutional analysis of the perceptual kind of a thing begins precisely with a study of "the passive transformation group." For the notion of passive and active transformation groups in physics, see Eugene Wigner, Symmetries and Reflection (Bloomington: Indiana UP, 1957), p. 45.

Digitized by Arya Samaj Foundation Chennai and eGangotri mental phenomenon can be prepared in a stable way by standard instrumental procedures so as to exhibit its symmetry structure. If it can, then it is a perceptual object in a world in which the equipment and laboratory skills for constituting such phenomena experimentally is available. I believe such a description applies to such things as electrons, positrons, muons, and possibly neutral currents and quark combinations, insofar as all of these can be constituted within traditions of laboratory practice and presented as stable phenomena. (Cf. the discussions of muons in Galison, 126–133, and the general discussion on pp. 252–262).

THE EXPERIMENTAL ROLE AND THE THEORETICAL ROLE

O is the object-for-knowing constituted by the experimental role (S_x) , say, Milliken's discretely charged oil droplets. M is the set of instruments used in the preparation-presentation process. I use Dirac notation $|X\rangle_x$ to represent the experimental profiles of the phenomenon O, because I do not want to imply that O is a classically objective space-time object, that is, I want to make room for the possibility that not all of its experimental profiles are realizable under arbitrary circumstances. When O is a phenomenological symmetry, i.e., a stable, reproducible, empirical object, then it has the structure of a Husserlian phenomenon.

How does one refer to the experimental profiles of a scientific phenomenon? They are distinguished from one another by the sets of measure numbers that substitute for the symbol 'X' in $|X\rangle_x$, e.g., $|X\rangle_x$ becomes $|q_1, q_2, \text{ etc.}\rangle_x$, where $q_1, q_2, \text{ etc.}$, are the measure numbers provided by measurements with M. The measure numbers denominate the profiles and distinguish among them. If this is what is meant by 'describing', then the measure numbers describe, but this, I believe, is not everything a description does; description also gives the sensory qualitative appearances or profiles of a thing.

⁶ Dirac notation highlights the transformation theory aspect of Husserl's theory of the phenomenon. In classical physics, the ket vectors |X⟩ would be replaced by a classical (n-dimensional) vector space, each vector representing a determinate objective state of the system. In a nonclassical theory such as the quantum theory, ket vectors belong to a Hilbert space, and only those have phenomenological meaning which are contextually linked to measurement.

The 'X' (in $|X\rangle_i$) will be replaced, after measurement, by some set of definite numbers, q_1 , q_2 , etc., standing for one complete set of characteristics for a given context of measurement. It is usually assumed that the profiles of a scientific of measurement. This is the source of the view—held by Husserl and many others —that measurement merely "idealizes" aspects of phenomena already there. I criticized such views in my "Husserl's Later Philosophy of Science"; cf. also my "Natural Science as a Hermeneutic of Instrumentation," Philosophy of Science, L, 3 (1983): 181–204. Cf. also Hacking, Representing, p. 240.

How do the phenomena appear to S_x ? The appearances of a phenomenon are a function of the (usually) standard measuring context M; this "dresses" the phenomenon O qualitatively with the characteristic set of profiles $|q_1, q_2, \text{etc.}\rangle_x$. Using the analogy with our bodily senses and sensory powers that "dress" everyday perceptual objects with color, texture, hardness, smell, etc., as well as size and shape, M in like manner "dresses" scientific objects as perceptual objects for our knowing in characteristic sets of sensory–technological "garb." The description of the phenomenon is both quantitative and qualitative.

The measure numbers q_1 , q_2 , etc., are obtained from a "reading" of the measuring instruments M (cf. Heelan, 206/7). What is the connection between the preparation-cum-presentation of the phenomenon O and the numbers "read" from the signals? In the first place, one may validly "read" numbers from the instruments only when the phenomenon O is present to S_x , for only then do the events (and numbers) signal the empirical profiles of a phenomenon O "dressed" by the equipment M (Galison, 73/4, 88). In the second place, having the right numbers does not guarantee the existence of a phenomenon, for there are other factors involved, such as the instruments, standard procedures, experimental skills, laboratory traditions, and the social context of the research community (Galison, 86-110). In the third place, having the phenomenon does not depend on having one (uniquely) right set of numbers, for different laboratory traditions may produce systematically divergent measures.8 Alternatively, the phenomenon could be the outcome of a different theoretical background (classical, say, rather than relativistic, giving a different set of characteristic numbers) and different instrumentation.9 Just as a musical score always goes together with performer, instrument, audience, and musical tradition, without which there can be no music, so a theory always goes together with experimenter, equipment, laboratory traditions, and the social

⁸ According to J. S. Hunter, "The National System of Measurements," *Science*, CCX (November 21, 1980): 869–874, for almost all measures other than the International System of Units, repeatability (within a single laboratory) and reproducibility (among different laboratories) is poorly estimated, implying the existence of different and virtually divergent laboratory traditions for many scientific measures. Some of these divergences usually taken to be random may conceal undisclosed systematic factors.

⁹ Consider the history of the theory of the electron (Galison, 31–52), or the history of the theory of the positron (Galison, 86–110), first as related to the "birth-cry of atoms," and then as due to Dirac pair-production, or the history of the theory of the muon (Galison, 126–131), first as the "red" electron that caused cosmic ray showers, and then as the cosmic "green electron" with singular penetrating power (cf. also the general discussion in Galison, 252–262).

Digitized by Arya Samaj Foundation Chennai and eGangotri context of science, without which there can be no scientific

phenomenon.

Although the measured values of measure numbers are often taken to be data, strictly speaking data are not—and should not be taken to be—just numbers, but phenomenological profiles $|X\rangle_x$. These are experimental state vectors individuated by their individual set of measure numbers. The true datum is the vector $|q_1, q_2,$ etc. \rangle_x . Much discussion of the relationship between theory and data, even of the more enlightened kind, such as Robert Ackermann's, is unsatisfying for want of a distinction among the following: (1) measure numbers, (2) the model state vector $|X\rangle_t$ specified by the numbers, and (3) the phenomenological state vector $|X\rangle_x$.

Let us take a closer look at theory. Every "reading" of the measuring instruments in terms of measure numbers supposes a theoretical model. I have chosen to represent such a model by a set of Dirac vectors $|X\rangle_t$, with its transformation groups defining its model invariants or symmetries. The theoretical vector $|X\rangle_t$ is linked with the empirical profile $|X\rangle_x$ through the events—"measurement events"—that occur in the measuring instruments (cf. Galison, 126–131).

Consider S_x . For S_x , the "measurement events" play a part in the constitution of the phenomenon-for- S_x . Measurement "dresses" the phenomenon O with a set of characteristic profiles. The "measurement events" are not profiles of O; they are no more profiles of O than, say, red patches are profiles of a strawberry, for a profile is experienced as nothing but the presence of a perceptual object under one of its profiles. A profile of a red strawberry is one particular view of a red strawberry; a red patch is something on its own (from which possibly one may infer the presence of a strawberry or something else), but it is not nothing but the presence of a strawberry. (How right are the arguments of those—Wilfred Sellars and Husserl—who reject the "Myth of the Given"!)

THEORY AS SEMANTICALLY DESCRIPTIVE ONLY IN SOCIAL CONTEXT Now consider S_t . For S_t these "measurement events" are potentially "readable" events, that is, they are signals, subject to the proviso that, at the time of the "reading," the phenomenon is present to S_x . The process of passing from the measurement event as signal (which is physical) to a number (which is nonphysical) is a species of *interpretation*; it is what we do when, for example, we "read" a numeral (a physical sign or signal) as a number (that which under the right circumstances the numeral is taken to signify). Measurement events are then not just plain events, like the red patches mentioned above, but they are the physical signifiers of theoretical vectors, those

Digitized by Arya Samaj Foundation Chennai and eGangotri which within the experimental situation are characterized by measure numbers. 10

From the standpoint of S_t , the measurement events qua signals comprise a semiotic system, that is, a system of differences which functions as such. Events comprise a semiotic system, not because of any features the events may have as physical events but because, as a system of differences, they can function as signifiers. Semiotic events may take many forms and still remain events of the same semiotic system (cf. Galison, 248/9). Signifiers, moreover, are timeless and ideal, and so they can be taken to refer to past or future phenomena. Semiotic events, then, partake of a certain physical arbitrariness; by social convention, however, they will have standardized or canonical shapes. What they signify—their semantic meaning—comes in part from the abstract theoretical model $|X\rangle_t$ and in part from the context of use. What they refer to is the phenomenon O under its set of profiles $|X\rangle_x$. The theoretical model $|X\rangle_t$ now appears from the standpoint of S, to be a kind of language under which the reference phenomenon with its empirical profiles |X>, is described. Such a relationship is hermeneutical in the typical epistemological sense of that term.

The syntax of that "language" is mathematical; of itself it provides no more than what Husserl would have called a formal ontology or a possible formal ontology, that is, empty schemata of categories of things.¹¹ Its semantics, however, is tied up with the standardized experimental praxis within which it is used. Semantically it needs a social-historical-technological context to complete its meaning, and this is an open context (Galison, 254). From the point of view of S_t , then, the following analogy holds: the theoretical model (e.g., electron_t) is to (the kind of) phenomenon (e.g., electron_x) as the linguistic model (the abstract signifier) is to the signified (the kind of thing the phenomenon is). So far I have been speaking of the research role S_t.

It is different for S_x. What is overlooked when one looks out from the standpoint of S, alone is the constitutive function that the measurement events play in the preparation-cum-presentation of the phenomenon O within the experimental role S_x. A semiotic sign or signal connotes a semantical meaning. It is not generally the case that a semiotic sign also functions to constitute its referent object as an

See, for example, Husserl, Formal and Transcendental Logic, Dorian Cairns,

trans. (The Hague: Nijhoff, 1969), p. 120.

¹⁰ For the nature of semiotic systems, see, e.g., Ferdinand de Saussure, Course in General Linguistics, Wade Baskin, trans. (New York: Philosophical Library, 1959), or textbooks in semiotics.

Digitized by Arya Samaj Foundation Chennai and eGangotri object-for-knowing. Measurement events do just this, for they belong to the process whereby S_x constitutes the phenomenon as present, knowable, and actually known. This process is also hermeneutical, not now in the epistemological sense, but in an artistic—"existential" or "ontological"—sense, as resulting from a species of experimental performance. Let me clarify this.

Consider, for example, the way a violinist addresses her instrument with the bow in order to produce a piece of music. The bow in this case is guided by a memory of the piece to be realized, its position on the strings interprets the needs of the memorized piece, and the sounds produced are in turn interpreted by the listening ear. The score is the theoretical "description" of the piece, it stands for $|X\rangle_i$; the score alone, however, is insufficient for a performance: to have music, one needs in addition an instrument, a performances fulfill different purposes: to teach, to rehearse, to celebrate, etc. Finally, every performance relates to some historical tradition of good performance.

But most important of all, for every good performer, the role of the score undergoes a transformation when it ceases to be a theory and becomes instead a mnemonic, then the artist's scorebook becomes a set of "places" or topoi, the function of which is to remind the artist of the suites and sequences to be performed. As such it is a local, personal, contextual, historical, technological, and artistic guide, it an open or endless set of memory cues, it is no longer a universal theoretical prescription. Such a mnemonic belongs to the tradition of (what Frances Yates calls) "the art of memory."

Just so in science, once a skilled experimenter knows how within a laboratory tradition to prepare and present a scientific phenomenon, theoretical models become converted into mnemonics for performance and demonstration. The last word in scientific research is then with S_x . Theory as such has a formal place only up to the moment when a laboratory tradition permits the realization and reproducibility of stable phenomena suitable for the research needs of the social-historical situation; then it can drop out (Galison, 244, 252–262; and Crease, 337–350) for it is now realized in praxis.

¹² Methodological or epistemological hermeneutics is the conscious and deliberate work of interpreting a text or a set of signs or symbols. Existential hermeneutics, introduced by Martin Heidegger, is that ontological character of all human understanding whereby it interprets life in terms of Being. For a useful survey, see 1980). Joseph Rouse's Knowledge and Power (Ithaca: Cornell, 1987) is an impornatural science.

When the last word is said, how does Sx use the language of theory? Substantive terms used in the theory, such as 'electron', etc., refer to phenomena, i.e., to electron, etc. Each kind of phenomenon is defined in terms of its characteristic set of profiles, e.g., of mass, charge, spin, lepton number, strangeness, etc.; these are the components of $|X\rangle_x$. In each category, the profiles are distinguished from one another by measure numbers that specify and distinguish the contextually realizable vectors of the theoretical model. Some of these measure numbers may be additive, some ordinal, and some merely serve to make discriminations. The descriptive qualities of the phenomenon-what I mean by the dress of the phenomenon-are given by the kinds of standard preparation devices or laboratory practices established by the research community. When the last word is said, the vocabulary of theory remains, but no longer as "theory-laden"; its meaning has become instead "praxis-laden" (cf. Galison, 255).

Whether or not the memory of its theoretical origins is retained, the phenomenon becomes—or seems to become—a part of nature. And though the old terms have their origins in theory and continue to be used, their semantic content changes, bypassing theory and signifying directly the characteristic empirical qualities of the "dressed" phenomenon. That "naturalized" artifact of human culture, the scientific phenomenon, is now, in the sense appropriate to the historical culture of that scientific community, a thing of nature. In our culture, electrons, positrons, muons, and possibly neutral currents are things of nature.

REALISM, ANTIREALISM, OR IDEALISM?

So much for the epistemological question. What now about the onto-logical question? If phenomena—i.e., perceptual objects—are real, then scientific phenomena are real. Then electron_x, positron_x, muon_x, etc., are real and belong to appropriate regional ontologies of the cultural—historical world in which we presently live. As for black holes, quarks, and such like, about which physicists are presently undecided, one will say that laboratory physicists will decide eventually whether these are—can be made to exhibit the phenomenological symmetries of—scientific phenomena (cf. Galison, 126–131). If they become phenomena, then they are—i.e., they become for us, for our culture—real. Reality, in this view, is naturalistic and evolutionary, but people and praxis are what provide it with the categories of the real. Within this picture, the objects of

¹³ Of those who defend a naturalistic evolutionary realism of this kind, some, such as C. S. Peirce, belong to the pragmatist tradition and some, such as Jürgen Habermas, to the Marxist tradition.

Digitized by Arya Samai Foundation Chennai and eGangotric entities like theory—electron, positron, etc.—are ideal semiotic entities like words, and, like words, they represent no more than empty schemata of an indeterminate and undecidable formal ontology outside of social-historical-technological contexts.

If, to the contrary, one takes formal ontologies to give categories of reality apart from the means of realizing or fulfilling the schemata they define, then theoretical entities such as electron, positron, etc., should be taken to be, not as in the first position the meanings of word-like semiotic entities, but to refer to real categories independently of the social-historical-technological context of experimental fulfillment. This Platonizing position is today called "Scientific Realism."14

There is a third or transcendental position that lies in between the two just mentioned and has certain attractive features. Husserl gave it preference in his own philosophy of science. In this view, theoretical models are not constitutive of an ontology—they are not the categorial contents of a realism-but they are nevertheless universally and transcendentally regulative for all lifeworlds, and they are exemplified in historical culture-relative ways. The fulfillment of such formal conditions then is not a contingent social-historical-technological event as it would be for the first position, but a necessary -though not a sufficient-condition governing all constitutions of phenomena and all constitutions of historical life-worlds. The necessity in question belonged, as Husserl thought, to the privileged status of scientific knowing.

My personal view tends to favor the first position. I have criticized Husserl's view in a paper cited above. I find that the reasons he gave are bad reasons. Whatever choice is made, it has to be made against a background that is much broader than that of science itself, an open background that involves history, culture, religion, and art, and the other ingredients of experimental performance so well described and analysed by-to mention just two-Galison and Hacking. 15

PATRICK A. HEELAN

State University of New York/Stony Brook

15 For an excellent analysis and critique of the current state of the question concerning truth and realism in science, see Joseph J. Kockelmans, "On the Problem of Truth in the Sciences," in Proceedings and Addresses of the American Philosophical Association, LXI, Suppl. (1987): 5-26.

¹⁴ Wilfred Sellars gave the now classical defense of scientific realism in "Philosophy and the Scientific Image of Man," in Science, Perception, and Reality (London: Routledge & Kegan Paul, 1963), pp. 1-40. Of particular note among the many recent works on this topic are Clifford Hooker, A Realistic Theory of Science (Albany: SUNY Press, 1987); Rom Harre, Varieties of Realism (New York: Blackwell, 1986); and Scientific Realism, Jarrett Leplin, ed. (Berkeley: California UP,

Digitized by Arya Samaj Foundation Chennai and eGangotri PHILOSOPHY IN THE LABORATORY*

LTHOUGH they are from different traditions, Ian Hacking and Patrick Heelan find common ground: they deplore the impoverished representation of experiment in modern philosophy of science. Both writers seek simultaneously to capture the rich historical diversity of experimentation revealed in recent historical studies, and the abstraction needed for philosophical inquiry. And both defend a thesis of phenomenal realism based on the stability of experimental results.

Hacking offers an insightful taxonomy of experimentation which I have reorganized from his twelve items (counting subcategories) into four topics: (1) the focus of experimental inquiry (e.g., choosing between rival theories); (2) the establishment of knowledge prior to experimentation (background knowledge, high theory, instrument knowledge, "topical hypotheses" that bind theory to experiment); (3) experimental materiel; and (4) data and data manipulation (data production, assessment, reduction, analysis, and interpretation). Heelan, too, appreciates the multiplicity of factors contributing to experimentation: "instruments, standard procedures, experimental skills, laboratory traditions, and the social context of the research community." Such taxonomies are needed to understand when and how theory shapes experiment.

Both Hacking and Heelan plea for a naturalized realism grounded in the laboratory: Hacking for an "entity realism," and Heelan for a realism about "scientific phenomena." To get there, each defends the robustness of entities. Building on Edmund Husserl, Heelan advances an intriguing metaphor: just as theoretical physicists search for symmetry groups (e.g., gauge theories) that leave certain theoretical terms unchanged, so, Heelan contends, experimentalists search for phenomena left standing under changes of (1) the state of the observer and (2) the state of the object under investigation. The exploration of these two "practical" symmetry groups and the con-

^{*} Abstract of a paper to be presented in an APA symposium on The Philosophical Significance of Experimentation, December 28, 1988, commenting on papers by Ian Hacking and Patrick A. Heelan, this JOURNAL, this issue, 507–514 and 515–524, respectively.

See, for example, Peter Achinstein and Owen Hannaway, eds., Observation, Experiment and Hypothesis in Modern Physical Science (Cambridge: MIT Press, 1985); Allan Franklin, The Neglect of Experiment (New York: Cambridge, 1987); my How Experiments End (Chicago: University Press, 1987); David Gooding et al., The Uses of Experiment (New York: Cambridge, forthcoming); Steven Shapin and Simon Schaffer, Leviathan and the Air-Pump (Princeton: University Press, 1985).

Digitized by Arya Samaj Foundation Chennai and eGangotti ditions of their implementation make up (as I understand it) "constitution analysis" applied to the experiment/theory relation. That there are such groups—and corresponding invariants—underwrites Heelan's argument for a realism about phenomena.

Hacking defends entity realism against Thomas Kuhn's different—worlds thesis by emphasizing the survivability of entities in a universe of shifting theories. In particular, Hacking stresses the malleability of each of his taxonomic groupings—especially topical hypotheses—as scientists strive to reconcile experiment and theory; but there remains a relative rigidity of entities. According to Hacking, this stability sanctions incommensurable truths about a stable ontology rather than incommensurable ontologies.

Stability, as I see it, must come from more than the convergence of experiment and theory. Logical positivists in the 1920s sought convergence by reducing all science to elementary observation "protocol" sentences. Positivist historians concurred that the proper periodization of science was on two levels: a continuous basis of aggregated observations, and a discontinuous sequence of theories variously organizing the "facts." When the postwar philosophical antipositivists came to prominence, they stood the old scheme on its head: Kuhn (among others) took theory as the ground and observation as the superstructure. Changes of theory dictated changes of observation.

Despite their apparent opposition, both positivists and antipositivists remained committed to the view that the stability of science hinged on whether or not experiment and theory could be reduced to one cohesive bundle. Both take for granted that there is one underlying narrative line (observation for positivists, theory for antipositivists) on which all else rests (epistemically and historically). What is needed is a heterogeneous representation of the periodization of modern science, allowing breaks in theory, instrumentation, and experimentation. The stability of the scientific enterprise rests (in this scheme) not on the total unification of science based on experimental or theoretical reductionism, but on the contingent fact that (1) there are traditions within experiment, theory, and instrumentation; (2) the dislocations within these "subcultures" of physics are not all synchronous; and (3) there are only piece—wise connections between the different strata, not a total convergence or reduction.²

² On experiment and theory as subcultures of physics, see my *How Experiments End*; on periodization and reduction see my "History, Philosophy, and the Central Metaphor," *Science in Context* (forthcoming, 1988).

The recent discovery of high-temperature superconductivity illustrates a domain of empirical inquiry which proceeded within a tradition of experimentation but against the tenets of accepted theory;³ the last decade of work on superstrings exemplifies a theoretical field strongly linked to traditions of quantum field theory and general relativity, but many orders of magnitude distant from connection to laboratory observation. Constraints other than the convergence of experiment and theory are at work (e.g., instrumental tests of superconductivity and theoretical constraints based on renormalizability and symmetries). These constraints on theory and instrumentation give rise to a stability and partial autonomy of these strata similar to that of the stratum of experimentation. Paradoxically, then, the stability of science may be better understood from its disunity than from an artificial unity imposed by the total reduction of science to observation, to high theory, or to experimentation.

PETER GALISON

Stanford University

³ On experimental traditions see: my "Bubble Chambers and the Experimental Workplace," in Achinstein and Hannaway; and Galison and Assmus, "Artificial Clouds, Real Particles," in Gooding, ed.

Digitized by Arya Samaj Foundation Chennai and eGangotri POSTMODERNISM AND RATIONALITY*

ODERNITY and rationality are interconnected in a conceptually compulsory manner in that both notions stand in an intrinsic and internal relation to one another. According to Jürgen Habermas, therefore, postmodernism, as a farewell to modernity and its discourse as a whole, would of necessity represent a departure from rationality. Postmodernism, he has argued, is in pursuit of an "immemorial" anarchism which, in the aftermath of Nietzsche, characterizes the Other of reason in essentially two forms -forgotten Being (Martin Heidegger) and ecstatic sovereignty (Georges Bataille). Such an approach, which claims the status of an esoteric, specialized knowledge (Sonderwissen) that is beyond discursive thinking and critical authority, would culminate in the business of deconstruction, a mode of thinking disconnected from all serious scientific analysis. This mode of thinking, moreover, can only result in an empty and formalistic incantation of an undetermined authority. Deconstructive thinking, according to Habermas, does not merely dismiss peremptorily the objection of pragmatic inconsistency; it makes such an objection altogether irrelevant by collapsing the distinction between literature and philosophy. Habermas remarks: "When, as Derrida recommends, philosophical thinking becomes released from its duty of solving problems and is turned into literary criticism, it is robbed not only of its seriousness, but of its productivity and efficiency as well" (246). From a perspective such as Habermas's, it thus makes sense to launch an investigation of the philosophical relation between postmodernism and rationality. But the outcome of such an investigation is exclusively polemical. In the history sketched out in The Philosophical Discourse of Modernity—a history characterized in terms of a play of reciprocal outdoing or outbidding (Überbietung, 74), in which one thinker tries to outwit his predecessor with increasing radicality until all the means of the paradigm that they are simultaneously criticizing and drawing on are exhausted—Habermas has nothing positive to say about the criticism to which reason has been subjected from the left Hegelians to Heidegger, Theodor Adorno, Jacques Derrida, and Michel Foucault. Habermas does, admittedly, seek to return to "the level of the

^{*}To be presented in an APA symposium on Postmodernism and Rationality, December 29, 1988. Charles Scott will comment; see this JOURNAL, this issue 539-540.

¹ Der philosophische Diskurs der Moderne (Frankfurt: Suhrkamp, 1985). [All the translations are mine. All page references in the text refer to this work.]

critique of reason inaugurated by Kant" (359), and "the traces of communicative reason" to be found in the early Hegel (43). His modest attempt, he claims, is capable, as he repeatedly stresses, of bringing about nothing less than a change of paradigm. Yet practically nothing in the history of thinking he outlines can be put to any use in his own attempt to undercut (unterbieten) this history of criticism. In a curious disregard of the practice of determined negation which he advocates so forcefully when it comes to criticizing others, Habermas concludes that postmodernity and rationality are simply incommensurable.

Habermas's abstract negation of the whole tradition of the modern questioning of reason becomes epitomized in his rejection of the idea of a comprehensive notion of reason. This is clearly directed toward Karl-Otto Apel's2 confidence that the different forms of the criticism of rationality-including even what he construes as the most complete neglect of the rational criterion of pragmatic consistency in French poststructuralism's effort to decenter the subjectare themselves rooted in "a more comprehensive form of rationality" (ibid., 19). By taking recourse to the old (Kantian) distinction between intellect (Verstand) and reason (Vernunft), Apel does indeed seek to understand the different sorts of criticism which have been directed against reason in the framework of the tradition of a philosophical self-limitation and self-differentiation of reason. The various types of this criticism would thus be fully justified, since they would be aimed against making absolute, abstractly isolated, and hence deficient forms of intellectual rationality (Verstandesrationalität). Rather than imply a parti pris for irrationality, such criticism of reason would consequently be rational, according to Apel. For, apart from the fact that it could not be developed without an at least tacit recognition of philosophical discursive rationality, such a criticism of reason would represent a matter of reason itself.

Habermas's rationale for refusing to give any serious credit to the tradition he is reviewing is that its criticism of subject-centered reason is self-defeating. It defeats itself, moreover, because, rather than subject such a notion of reason to an immanent critique, it radicalizes that criticism. That is, it displays a criticism that opens an attack on the (rational) roots of criticism itself. By challenging the very foundation of criticism, such a criticism of reason has, for Habermas, become total and self-referential (selbstbezüglich). Furthermore, he

² "Das Problem einer philosophischen Theorie der Rationalitätstypen," in Rationalität: Philosophische Beiträge, Herbert Schnädelbach, hrsg. (Frankfurt: Suhrkamp, 1984).

Digitized by Arya Samaj Foundation Chennai and eGangotri notes that such a totalizing self-referential putting-into-question of reason becomes entangled in the classical aporias or paradoxes of the philosophy of subjectivity, because it must make recourse to the very tools of subject-centered reason to debunk its authority. Even a radical criticism of reason which would seek to avoid the paradox of self-referentiality by reverting to the realm of rhetoric-as a de Maneanly alienated Derrida would do-would not escape it, because the merely inverted (subjectivist) foundationalism of such an attempt would not only fall prey to the weaknesses of the criticism of metaphysics all over again, it would also represent an ultimate "deadening of the blade of the critique of reason" (246). Such entanglement in the paradoxes of self-referentiality is inevitable, Habermas contends, if one continues to fight, as do Heidegger, Adorno, and Derrida, the "'strong' concepts of theory, truth and system," that is, the universal status that philosophy once claimed for the answers to its universalist questions. Yet reason, according to Habermas, is not bound to hold on to "these classical goals that metaphysics aimed at from Parmenides to Hegel" (246). Since Hegel, moreover, these traditional philosophical issues no longer exist. It would therefore be superfluous for any criticism of reason "to grasp the roots so deeply that it barely escapes the paradox of self-referentiality" (246/7).

It is not possible in the space I have here to show that Habermas does not do justice to the philosophers he discusses in his book. Therefore, I want to shift the terrain of the discussion. First, I wish to claim that the postmodern interrogation of reason which has triggered a renewed interest in the question of rationality in philosophy is not exhausted by the rediscovery of the irrational, whether as a boundary problem of rational cognition or as what puts the foundation of Western culture into question. From the Greek mathematicians who discovered the irrational as a limit of rationality, through the moment when the irrational becomes a philosophical problem properly—that is, when philosophy puts the subject in the very center of its concerns (and the irrational becomes thus determined as a region inaccessible to human cognition)—the irrational has been the unavoidable counter concept of rationality.3 It is precisely this conceptual machinery that is put into question in certain of the works that are labeled postmodern. Jean-François Lyotard4 can claim that postmodernism corresponds to the occurrence of a "major shift in the notion of reason," and thus in the very nature of

⁴ The Postmodern Condition: A Report on Knowledge, Geoff Bennington and Brian Massumi, trans. (Minneapolis: Minnesota UP, 1984).

³ Historisches Wörterbuch der Philosophie, Joachim Ritter, hrsg. (Darmstadt: Wissenschaftliche Buchgesellschaft, 1971ff), vol. IV, p. 583.

what we understand by knowledge, it is indeed because of what he believes to be the role of paralogism in contemporary thought, whether scientific or artistic (*ibid.*, p. 43).

Let me first recall that, for Lyotard, one ought to conceive of the postmodern as constituted primarily by a crisis of narratives, more precisely of the grand narratives such as the speculative narrative or the narrative of emancipation, whose function it had been to unify or legitimize both knowledge and social practice. As a result, this postmodern incredulity toward metanarratives, and the subsequent recasting of the function of legitimation, leads to an explosion of mini narratives or language games which profoundly affect the status of knowledge and the idea of reason. Lyotard exemplifies this transformation of knowledge with respect to the language games of science and its new practices of validation. He argues that science—and, in particular, scientific research and argumentation—puts all its stakes in working on proofs, rather than be oriented by a quest for performance. Yet what this means is that the pragmatics of science is essentially geared toward "searching for and 'inventing' counter-examples, in other words, the unintelligible; supporting an argument means looking for a 'paradox' and legitimating it with new rules in the games of reasoning" (ibid., p. 54). Indeed, the major characteristic of modern science, according to Lyotard, is that it "emphasizes the invention of new 'moves' ['coups' nouveaux] and even new rules for language games" (ibid., p. 53). This notion of "moves," which is quite decisive for what Lyotard claims to be going on in contemporary science, describes the agonistic dimension of knowledge formation. Scientific knowledge becomes plural and its languages become manifold to such an extent indeed that science, while playing its own game, is not only incapable of legitimating other language games, "it is above all incapable of legitimating itself, as speculation assumed it could" (ibid., p. 40). But, as Lyotard insists, this does not mean that the question of legitimation would become obsolete altogether. On the contrary, the most "striking feature of postmodern scientific knowledge is that the discourse on the rules that validate it is (explicitly) immanent to it" (ibid., p. 54). "For," he adds, "it is not philosophy that asks this question of science, but science that asks it of itself' (ibid.). But such inclusion "within scientific discourse [of] the discourse on the validation of statements held to be laws" gives rise, of course, "to 'paradoxes' that are taken extremely seriously and to 'limitations' on the scope of knowledge that are, in fact, changes in its nature" (ibid., pp. 54/5). As a result, rather than have a universal metalanguage that would legitimize and systematize all pieces of knowledge, the myriad of new moves is accompanied by another

Digitized by Arya Samaj Foundation Chennai and eGangotri myriad of newly invented rules for the arguments advanced. Lyotard characterizes the shift in the notion of reason as follows:

The principle of a universal metalanguage is replaced by the principle of a plurality of formal and axiomatic systems capable of arguing the truth of denotative statements; these systems are described by a metalanguage that is universal but not consistent. What used to pass as paradox, and even paralogism, in the knowledge of classical and modern science can, in certain of these systems, acquire a new force of conviction and win the acceptance of the community of experts (*ibid.*, pp. 43/4).

But not only these formal characteristics explain the postmodern understanding of knowledge and reason. As Lyotard tries to show, modern science is thematically geared toward new objects as well, in particular toward singularities and incommensurabilities. Lyotard concludes that what one learns from this research

is that the continuous differentiable function is losing its preeminence as a paradigm of knowledge and prediction. Postmodern science—by concerning itself with such things as undecidables, the limits of precise control, conflicts characterized by incomplete information, 'fracta', catastrophes, and pragmatic paradoxes—is theorizing its own evolution as discontinuous, catastrophic, nonrectifiable, and paradoxical. It is changing the meaning of the word knowledge, while expressing how such a change can take place. It is producing not the known, but the unknown (ibid., p. 60).

What thus seems to distinguish postmodern reason from classical or modern reason is its irreducible plurality and its willingness to sustain it. As Lyotard insinuates at the beginning of *The Postmodern Condition*, postmodern knowledge "refines our sensitivity to differences and reinforces our ability to tolerate the incommensurable" (*ibid.*, p. xxv).

What this brief account of Lyotard's analysis of the shift in the notion of reason in postmodern science indicates is that postmodernity is not a farewell to reason altogether, but a shift toward the reason of the plural, the indeterminate, the random, the irregular, the formless, the paralogistic. It is with respect to these that the unified, the lawful, the regular, etc., are shown to be derivative formations. What this account also shows is that the concern with rationality in our present time is perhaps more than just an attempt by philosophy to react retrospectively to discussions and events that did not take place within the region of its core. Herbert Schnädelbach⁵ has argued—and I follow his lead—that, if rationality is thematized today by philosophy, it is because philosophy seeks, by means of a detour, to reappropriate a traditional, if not hereditary,

^{5 &}quot;Einleitung," in Rationalität.

theme in a transformed shape. Indeed, although the concepts of rationality and reason have the same linguistic origin, they clearly are not substitutes for one another. Whereas rationality suggests solid scientific seriousness, reason is open to the suspicion of mentalism and metaphysical obscurantism. But reason in modern philosophy was never merely one faculty of consciousness among others. As Schnädelbach has argued, from Descartes to Kant and Hegel, "the philosophy of reason appears with the claim to secure the self-foundation of philosophy by reverting to the very reason of the philosophizing subject" (ibid., p. 8). Let me add that, if we look at reason in philosophy from antiquity on, reason, logos, appears as that which is constituted by the logon didonai, i.e., by giving an account while at the same time the philosophizing subject publicly accounts for itself. The concern with rationality and, in particular, with what happens to it in the advanced sciences is thus an attempt to reach beyond a notion of rationality defined by logical consistency to what Charles Taylor⁶ has called the "richer" concept of rationality and, ultimately, to philosophy itself. What is striven for here, however, is not a philosophy as it was before rationality superseded reason, but a philosophy in a transformed shape.

Nonetheless, it is questionable whether Lyotard's account of the transformation of reason within postmodernity allows such a philosophy to transpire in its most consequential and uncompromising form. Indeed, Lyotard's description of the new paradigm of reason sounds very much like an inversion of the old or modern one. The categories that he uses to pinpoint postmodern reason are the binary opposites of the traditional objects and values of reason. Discontinuity, undecidability, instability, and dissension cannot avoid drawing their very meaning from the horizon and telos of continuity, decidability, stability, and consensus. Lyotard's emphasis on paralogism and paradox as the postmodern form of legitimation—that is, on those operations of thought which Habermas abhors to such a degree that he appears unable to confront them in The Philosophical Discourse of Modernity-makes sense only with respect to the standards of logical and articulated argumentation. Let us therefore consider the following: if the postmodern transformation of the idea of reason consisted only in such value inversions and the subsequent dependency of what Lyotard calls "islands of determinism" on these inverted values, then it would be hard to understand all the fuss made over the so-called resurgence of irrationalism. Indeed, a perhaps much more decisive transformation of thought is in the offing,

⁶ "Rationality," in *Rationality and Relativism*, Martin Hollis and Steven Lukes, eds. (New York: Blackwell, 1982), p. 88.

⁷ Op. cit., p. 59.

which has much greater implications but which, at the same time, is also much less flashy than the usual value inversions.

What I am referring to is what I perceive to be the thrust of the thinking of Derrida, whom Habermas deigns to call a nicht argumentationsfreudiger Philosoph, that is, a philosopher with little enthusiasm for argumentation.8 Derrida, however, would certainly resist any association with what Lacoue-Labarthe has recently called the rag-bag nature of the concept of postmodernism; and for reasons not unlike those which made Heidegger refuse all identification of thinking with passing fads. Yet, if I nonetheless decide to discuss Derrida's thinking in this context, it is not simply because Habermas has reserved a very noticeable place for Derrida in his review of the postmodern debate with subject-centered reason. It is because Derrida himself has, on more than one occasion, spoken of his research as being concerned with the origin of reason, logic, and thinking. But, as we shall see, such a concern with origins does not mean, as Habermas believes, that the origins of reasons which Derrida reveals would consist of "contrasting concepts to reason . . . resistant to all attempts of rational incorporation" (127). Indeed, what appear in Derrida to be the origins of reason are not of the order of what Habermas understands as the Other of reason. Any attempt that explores the origins of reason, logic, and thinking cannot make an unconditional use of what it seeks to explain and account for. Origins cannot be described in the categories constitutive of what they make possible. This difficulty is, of course, all the more severe when it comes to thinking the origin of reason, logic, and thinking. The investigation of what is prior to them must, therefore, follow complex strategies and calculated conceptual games which are not entirely reasonable and logical compared to the standards that these origins are to render intelligible. This explains also why the origins of reason, logic, and thinking are not to be determined as the Other of reason, the opposite of reason, the irrational. These are categories of reason which designate, as Derrida puts it, "an opacity within the system of rationality." In Of Grammatology, Derrida decides to determine the origin of reason as "nonrational as the origin of reason must be." That this 'non' is not to be understood as a negation in

⁹ Of Grammatology, Gayatri C. Spivak, trans. (Baltimore: Johns Hopkins, 1976), p. 259.

⁸ This accusation is simply a convenient "argument" for evading an intellectual challenge, much as is the charge of free play. Whereas the first accusation is symptomatic of an inability to track the traditional modes of argumentation to the very displacement to which they become subjected in Derrida's writings, the incrimination of free play does not face the fact that truly patternless, truly arbitrary, truly random play is a highly complicated agenda only recently raveled out by some mathematicians.

a logical sense is obvious from Derrida's discussion, in the same work, of the "rationality" of such an origin. Rationality here does not mean conformity to reason, for Derrida notes that the word in question should "perhaps be abandoned," since what governs these origins of reason "no longer issues from a logos" (ibid., p. 10). This is indeed an excellent example of a calculated inconsistency and contradictoriness that becomes unavoidable—and for compelling philosophical reasons—if the "reason" of reason is to be interrogated. But what this example also suggests is that, for Derrida, such inconsistency, or paradox, is not a value in itself. It is a function of a language game aimed at determining the origin of reason according to standards and requirements that are intrinsically philosophical.

What is more, these nonrational origins of reason, which are "not yet logic but the origin of logic,"10 "the pre-logical possibilities of logic,"11 cannot be origins in a strict sense. The ground of the ground, as Heidegger has shown, cannot be another deeper, more excellent ground. Likewise, the "reason" of reason, the origin of reason, cannot be said to be a more profound origin. But, in spite of this similarity, there is a world of difference between what Heidegger circumscribes as the ground of the ground, and what Derrida designates as the origin of reason or, more simply, as the origin of the origin. Whereas for Heidegger the ground of the ground is Beingthat is, Being as the opening-lighting of beings and, hence, of the grounding relation that can only be thought between beings—the nonrational origin of reason is not to be conceived in phenomenological terms as an essential opening, a lighting for appearing. The nonrational origin of reason thematized by Derrida does not stand in a necessary relation to reason, as does the Heideggerian ground of the ground to what is called ground in philosophical thinking. As Derrida argues in Of Grammatology, the origin of reason is something that adds itself to reason; it is a supplement of reason, something exterior, nonnecessary to reason, but also that without which reason could not be what it is. These nonrational origins of reason are of the order of unheard-of trivialities. They are trivialities in the sense that Husserl gives to this term in Logical Investigations; they are of the order of what Heidegger calls the simple and are therefore all that more difficult to think. But the origins of reason to which Derrida points do not have the Husserlian characteristic of being essential forms and laws (of meaning) or the unitary character of

UP, 1986), p. 242.

"I Derrida, "Limited Inc abc," Samuel Weber, trans., in Glyph 2 (Baltimore: Johns Hopkins, 1977), p. 235.

Derrida, Glas, John P. Leavey and Richard Rand, trans. (Lincoln: Nebraska

Heideggerian *Einfachheit* and *Einfalt*. They are simply not subsumable within rationality. This is not because reason is powerless to think its nonrational origin, but, rather, because reason would not be what it is if it did. As Derrida puts it, reason "is constituted by that lack of power. It is the principle of identity." In the present space, I cannot do more than gesture toward supplementarity, arche-trace, differance, iterability, and the remark, and toward the rich contexts in which Derrida develops these notions as articulating the limits whence reason comes into its own. These origins of reason are incommensurate with reason, yet not as the Other of reason, but as the structures against which reason stands out as that which precisely has no Other, no exterior, except that which it lets unfold from within itself.

From everything that I have established until now, it should already be obvious that such a nonnecessary and nonessential relation between the origins of reason and reason itself does not allow one to speak of this relation, as Habermas does, as one of "an anonymous, history-causing productivity" (210). Habermas must, of course, make this mistake, just as he must misinterpret Heidegger's notion of Being as a counter-force (127). Otherwise, he could not write the history of the criticism of reason in terms of a history that never escapes "the foundationalist steadfastness of the philosophy of subjectivity" (211) in spite of all attempts to overthrow it by inverting foundationalism. Yet Being is not a force, just as the nonrational origins of reason do not "produce." Only reason produces, but its production is not without the nonoriginary structures at the origin of reason. Because of this nonnecessary and nonessential (yet not, for that matter, accidental or contingent, but, rather, inevitable) dependency of reason on its origins, they are not principles, conditions of possibility, or "radical" foundations either. For all these rest precisely on the origins in question. Hence, one cannot use the philosophical language of grounding, constituting, or opening to conceptualize the relation between the nonoriginary origins of reason and reason itself. These difficulties have misled Habermas to conclude that,

if the Other of reason is more than the irrational or the unknown, if it is, in fact, the *incommensurable* that reason cannot touch, save at the price of the explosion of the reasonable subject, then there are no conditions under which a theory that would reach beyond the horizon of what is accessible to reason and that would thematize, let alone analyze, the interaction of reason with a transcendent originary power, could be represented as possible in a meaningful way (277).

¹² Of Grammatology, p. 149.

Digitized by Arya Samaj Foundation Chennai and eGangotri From what I have said thus far, it should be evident that the terms in which Habermas formulates the problem do not pertain to what is under discussion here—the origin of reason is neither transcendent nor originary, nor is it a power. Moreover, if theory, thematization, and analysis are understood in the sense in which they have always been understood, that is, as thinking from and toward a whole, then no theory could ultimately thematize and analyze these nonoriginary structures of reason. Yet, as has already been indicated, these origins have a certain "rationality" about them which can be brought into view, and which within the parameters of a very calculated language game can be articulated with rather great precision. The same can be said of the relation that the nonnecessary origins of reason entertain with reason.

But, undoubtedly, another task remains, the task of discussing the status of these nonnecessary origins of reason with respect to and within the philosophical disciplines that traditionally have served to thematize questions of origins, principles, or conditions. One would in this regard have to follow up on Derrida's suggestion, in *Of Grammatology*, to develop a new transcendental aesthetics even if, as he notes, this could no longer be an aesthetics in either a Kantian or Husserlian sense. And, I would add, one ought to develop a new doctrine of the categories in order to come to grips with these non-originary structures of reason, although these structures are not forms of judgment, strictly speaking.

To conclude, let me return to Habermas's contention that the postmodern critics of rationality confuse the universalist questions that philosophy continues to honor with the claims that it traditionally has made about the status of its answers to these questions. In this context, Habermas distinguishes questions regarding the rationality of utterances and the conditions of their range as examples of such universalist questions. Although philosophy still understands itself as the guardian of rationality, and thus requests that all responses to universalist questions mirror themselves in the grammatical forms of universal utterances, philosophy must, he adds, also "calculate with the trivial possibility, that tomorrow or at some other place they might become revised" (247). Hence his plea for utterances with weak status claims. This trivial possibility that all utterances regarding universalist problems can be subject to revision is indeed for Habermas a trivial matter, and thus requires only a pragmatic response in terms of fallibilism. Yet, if this possibility can always happen to a strong utterance, then this is a necessary possibility that has to be accounted for in a manner that may affect the universal questioning of philosophy itself. Habermas sees no theoretical problem in answering universalist questions in spite of the

Digitized by Arya Samai Foundation Chennal and Gangotillosophy has claimed with respect to these truths. But, if the trivial possibility Habermas refers to is a necessary one, then everything that the criticism of rationality from Heidegger and Adorno to Derrida has directed against the strong concepts of theory, truth, and system will also weigh against the weak claim of utterances regarding their irrefutability. More importantly, Habermas has even fewer problems with the question in general, with the question's questioning form, a form that is invariably universal in thrust. What if precisely that form was at stake in certain of the philosophical writings that have been labeled "postmodern"? This form is at issue, not in order to give in to irrationalist modes of coping with philosophical issues, but, rather, to conceive of what in the logos is prior to its questioning form. Obviously, this something cannot be arraigned by a question: it cannot be thought if thinking, in its very essence, is constituted by the form of the question. 13 Still, as I suggested, within the limits of a certain very calculated language game, thinking's prequestioning forms can also find their articulation. Needless to say, within the discourse of philosophy, that is, within that thinking that the nonquestion "makes possible," any "thematization" of what is prior to the question must inevitably appear (to someone like Habermas, at least) as nonanalytical and stylistic criticism (223). Yet, if rationality is the issue today, is it merely to conjure the alleged lack of philosophical argumentation and discursive consistency? Is it not, rather, because thinking has discovered a new sort of finitude which requires not the abandonment of the traditional forms and claims that constituted it, but their displacement within operations of thought whose calculated economy obeys a "rationality" of its own? To pose the question of rationality with regard to the postmodern is then, perhaps, to recover a sort of thinking which, although it presupposes a displacement of all the essential forms constitutive of thought in the tradition of thinking, is more of the order of the philosophies concerned with the "richer" concept of reason and is thus in greater proximity to the traditional exigencies of thinking, both thematic and methodological, than all the celebrations of rationality, whether grounded in reflexive consciousness or in pragmatic consensus for-

State University of New York/Buffalo

RODOLPHE GASCHÉ

¹⁸ Glas, p. 191.

POSTMODERNISM AND RATIONALITY

Digitized by Arya Samaj Foundation Chennai and eGangotri THE QUESTION OF REASON*

OR Rodolphe Gasché, rationality is historically determined, and, if we are to treat postmodern thought intelligently-indeed rationally—we must invest our thought in the reflective and critical processes and procedures that have led to and formed postmodernism. Otherwise, we can do no more than irrationally condemn or follow it. Since postmodernism involves a departure from that type of rationality which we call modern, we must know well the modern texts, arguments, and vocabulary, the modern readings of previous texts and refutations of previous arguments, the emergence of modern thought, its crises and blindnesses, its exclusions and suppressions. We must know thoroughly the tradition, with its institutions and standards of behavior, from which postmodernism arises and takes its departure. Otherwise, we do not engage postmodern thought, communicate with it, or think it through. Gasché does not want to undercut the history of postmodern criticism of rationality or to undercut the postmodern critique in the way that Jürgen Habermas does, but to think in and through both the modern history of rationality and its critique, to think out of their effects and accomplishments. Hence, he retrieves and retains the questioned traditions of rationality in the postmodern overcoming of them. This procedure on Gasche's part is specific and concrete, and is itself a type of rationality, in contrast to the abstractness of Habermas's "negation of the whole tradition of the modern questioning of reason." One of the admirable aspects of his paper is his showing that Habermas's rejection involves less reason than Habermas's own position calls for and that it falls prey, if not to irrationality, certainly to a rationally truncated type of criticism which passes over too quickly its own history as well as that of postmodern thought. Habermas's often gross misreading of the texts that he criticizes is not incidental to the rationality that he employs.

After noting Gasché's use of Jean-François Lyotard and Jacques Derrida in accounting for postmodern rationality, we note that the major issue for postmodern thought is to carry out the transforma-

0022-362X/88/8510/0539\$00.50

© 1988 The Journal of Philosophy, Inc.

^{*} Abstract of a paper to be presented in an APA symposium on Postmodernism and Rationality, December 29, 1988, commenting on Rodolphe Gasché's paper of the same title, this JOURNAL, this issue, 528–538.

Digitized by Arya Samaj Foundation Chennai and eGangotri tion of reason in the way in which the postmodern thinker gives accounts of this transformation. We describe Gasché's own agenda and compare and contrast it to postmodern strategies for carrying out the transformation of reason. To what extent are his commentary on postmodern rationality, and mine on his paper, postmodern texts? How are we to describe postmodern texts without reinstating the rationality they are designed to overcome?

CHARLES SCOTT

Vanderbilt University

PROLEPTIC READING OF SOCRATIC DIALOGUES Digitized by Arya Samaj Foundation Chennai and eGangotri

PLATO'S CHARMIDES AND THE PROLEPTIC READING OF SOCRATIC DIALOGUES*

N several recent papers I have argued for a proleptic reading of what I call the premiddle dialogues: Laches, Charmides, Lysis, Euthydemus, Protagoras, Euthyphro, and Meno. I want here to apply this approach to the Charmides and to delineate more carefully the notion of prolepsis.

Prolepsis means the anticipation of things to come, and in this case the anticipation of philosophic views expressed in the middle dialogues (Symposium, Phaedo, Republic). But in what sense does Plato in one dialogue look forward to views he will formulate in a later work? For us with the whole Platonic corpus before us, it is possible to follow themes from dialogue to dialogue, looking forward and backward. For the original reader, no forward reference is possible. Now it seems safe to assume that the Charmides was written before the Republic. If, as I propose, we understand by the meaning of a dialogue the philosophical message that Plato may be thought to convey to his original readers, it cannot be right to propose a doctrine of the Republic as the true meaning of a passage in the Charmides. The meaning of a Platonic text must be something that is in principle accessible to the original reader. How then can such a dialogue be proleptic?

I want to suggest that one major function of the premiddle dialogues is to prepare the reader for the reception and understanding of later doctrines. A puzzling passage in an earlier work may be designed to provoke the reader into thinking about a problem to which Plato will later offer a solution. Hence, we as modern readers may legitimately use whatever we find in other dialogues as a possible clue to Plato's intentions in a given text, but no such hypothesis can be regarded as confirmed for a given text unless we find in the text in

* To be presented in an APA symposium on Plato and the Proleptic Reading of Socratic Dialogues, December 29, 1988. Charles L. Griswold, Jr. will comment; see this JOURNAL, this issue, 550/1.

These seven dialogues are my Group II, which I imagine to have been written in 386-380, when Plato was establishing his school in the Academy. (Hence the inappropriateness of calling them "Socratic" dialogues.) These are to be distinguished from Group I (Apology, Crito, Ion, Hippias Minor, Gorgias and Menexenus) which I suppose to have been written earlier. See my "Did Plato Write Socratic Dialogues?" Classical Quarterly, xxxi, 2 (1981): 305-320; "Plato's Methodology in Laches," Revue Internationale de Philosophie, xL, 1-2 (1986): 7-21; "Plato and Socrates in the *Protagoras*," *Méthexis* (forthcoming). For the earlier date of Group I, see "On the Relative Date of the *Gorgias* and *Protagoras*," *Oxford* Studies in Ancient Philosophy (forthcoming).

0022-362X/88/8510/0541\$00.90

© 1988 The Journal of Philosophy, Inc.

question some indication that the original reader might plausibly understand in this sense. Let us take this as a test for meaningful

prolepsis.

Consider two examples. Two themes from the Charmides which are destined to reappear in the Republic are (1) a definition of temperance as "doing one's own thing" (ta heautou prattein), and (2) a reference to "the knowledge of good and bad" as the source of a happy life. If we apply our test to these two examples, the results are rather different. In the first case, we as modern readers cannot fail to see an interesting connection between "doing one's own thing" as a definition of temperance at Charmides 161B and the same formula applied to justice in Republic IV: the former is used to forbid, the latter to enjoin the division of labor. (Compare Charmides 161E-162A with Republic II, 369B-372A; IV, 433A-D.) We find Plato toying in the Charmides with an idea that he will put to work seriously in the Republic. But there seems to be nothing in the text of the Charmides to serve as a clear signal to the reader of Plato's more serious intentions.2 In the case of the knowledge of good and bad, I shall try to show that the situation is different, and that we do have an example of meaningful prolepsis. This will be all the clearer if we take the Charmides not in isolation but as member of a group that includes Laches, Lysis, and Euthydemus. For it is part of my hypothesis that these four works were designed to be read together, and that the full meaning, the message to be conveyed, is a function of the group as a whole.

The Charmides and the Laches both conclude with an apparent failure to find the required definition, and yet in both cases Plato has presented, shortly before the end, a possible definition of virtue or temperance as the knowledge of good and evil. The thoughtful reader will see that we are in fact offered a positive definition in both dialogues, and that this form of knowledge has as its specific function to benefit us and to make our lives happy and successful (eu prattein at Charmides 174B11). Furthermore, toward the end of the Charmides Socrates begins to make explicit what is implicit in the choice of interlocutors for both dialogues: that the success in question is a matter of public welfare as well as private happiness. We are given

² Actually there is such a signal, but its correct interpretation seems to require some knowledge of the doctrine expounded in Republic II and IV. At 161C Socrates emphasizes that the meaning of ta heautou prattein is mysterious, and that its literal sense, forbidding the division of labor, could not be the nomos of a well-governed city (161E). And see below on the political theme in Charmides, Lysis, and Euthydemus.

the imaginary sketch of a society in which something like the Socratic elenchus is in charge. "We who possess temperance [understood here as knowing what you know and what you do not know] would live our lives infallibly, and so would those who are ruled by us" (171D, emphasis added). "The household managed by temperance would be well managed, and the city would likewise govern its affairs well, and similarly for everything that is ruled by temperance" (171E). But a moment later Socrates reports a dream (173A) in which he catches sight of the uselessness of temperance as previously defined, namely, as the knowledge of what one knows and what one does not know. Living and acting according to knowledge will be of no use at all unless it is knowledge of the appropriate sort. But what knowledge is responsible for a happy life? Of course it is the knowledge of good and bad. "Without this, we will lose the result that all things happen well and beneficially" (174C).

The political dimension opened up by this passage in the *Charmides* is pursued in a mock-serious vein in *Lysis* 207E–209C, where Socrates points out to the boy Lysis that his parents, who desire his happiness, nevertheless allow even servants to rule over him in those matters in which he lacks knowledge, but that "on the day that your father thinks you have better understanding (*beltion phronein*) than he does, he will entrust both himself and his property to you" (209C). Not only his father but all the Athenians and even the Persian King will do the same. "Greeks and barbarians, men and women, all will turn their affairs over to us, in those matters in which we become wise" (210A9–B2).

The playful tone of this passage from the Lysis might discourage us from taking it seriously if it did not follow so closely the line of reasoning just quoted from the Charmides, where happiness, benefit, and power are the consequences of wisdom, specified as the knowledge of good and evil. Our intelligent reader who puts together Socrates's dream in the Charmides with the Great-King-and-Athenians passage in the Lysis will see that Plato is articulating a vision of the political realm as a realm within which moral wisdom will rule. The reader so prepared should be able to make sense of what is otherwise a most puzzling passage in the Euthydemus where the royal art is introduced as producing happiness (291B), on the grounds that it is the only art that knows how to make good use of the products of the other arts (291C). The royal or political art will be the cause of right action (orthos prattein) in the city, "steering all things and ruling over all" (291C-D).

Now, this passage in the Euthydemus is explicitly marked as a kind of riddle. We have just been told by Socrates's young interlocutor

Cleinias that, as generals turn their victories over to politicians who know what use to make of them, so mathematicians, if they have any sense, turn their discoveries over to dialecticians to put them to use (290C–D). At this point, Socrates's narrative of the conversation with Cleinias is interrupted by Crito, Socrates's interlocutor in the frame dialogue, who doubts that Cleinias can really have said that: Socrates is not sure who said it. Perhaps he heard it from some superior being, i.e., from some divinity (291A). After this interruption the conversation with Cleinias is resumed, and we are given no further clue as to the mystery.

For readers of Republic v-VII this riddle is easily solved. The parallel between politikoi and dialektikoi in the description of the kingly art at Euthydemus 290C-D points to the doctrine of philosopherkings and to the role that dialectic plays in their education; whereas the subordinate position of mathematics is-for us-an unmistakable mark of Plato's own educational scheme. But this solution is unavailable to the interlocutors in the Euthydemus, and presumably it was equally unavailable to the first readers of that dialogue. Our hypothetical original reader, equipped with a good knowledge of other "Socratic" dialogues, will be able to see only that Plato is pursuing here the theme of political rule for wisdom which we have traced in Lysis and Charmides. He will recognize in the royal art of the Euthydemus an equivalent of the knowledge of good and evil in the Charmides, since both forms of knowledge have the same function: to make things useful, to produce happiness and correct action in the city (cf. Euthydemus 290D, 291C). And he will be prepared to recognize dialectic as somehow parallel to these two sorts of knowledge. But he will have no clear idea of what this knowledge consists in, or how it is related to dialectic. For that he would need to know something of the theory of Forms, of which there is no trace in the dialogues considered so far.3

Furthermore, if our specimen reader continues on in the *Euthydemus*, he will see that Plato has deliberately constructed this passage not only as a mystery but also as a source of paradox and puzzlement. For (it is argued) the kingly art must of course be beneficial and hence must provide us with something good. But earlier in the dialogue we agreed that the only good thing was knowledge (292B1–2, referring back to 281E). So the beneficial function of the political art will be to make us wise and good. But in what knowledge are we to become wise? And how is it to be useful? The only knowledge it could

³ For a partial exception, see below on the proton philon in the Lysis.

provide would be itself, or knowledge of itself. But, if its utility is to make others good, we fall into a regress: good for what? Good for making others good? And so on, without any hope of resolution. So we simply do not know what knowledge it is that can make us happy (292E).

Once again we modern readers can look ahead to Republic VI to see how this regress will be broken: by the highest object of knowledge identified as the Form of the Good. I see no hint of this solution in Charmides or Euthydemus. In Lysis, however, we must take account of a comparable regress. If everything dear is dear for the sake of something—A dear for the sake of B, B dear for the sake of C, etc.—"then necessarily either we wear ourselves out proceeding in this way or we will come to some first principle $(arch\bar{e})$, for the sake of which we say all other things are dear. . . . (They are) as it were images of it. . . . that first thing, that which is truly dear" (219C-D). From the Lysis alone we could never guess that it is the Form of the Good (or, in the context of the parallel at Symposium 210E-212A, the Form of the Beautiful) which is here being described. Again, it seems that only the modern reader, looking backwards from the Republic or the Symposium, is in a position to see this passage in the Lysis as proleptic.

But perhaps that conclusion would be too pessimistic, and our model reader of the *Lysis* and *Euthydemus*, if he is clever enough to put together the regress from one dialogue with that from the other, may be able after all to see that the move that blocks the regress in the *Lysis* ("dear for the sake of what?") can also serve to block the regress of the *Euthydemus*: "good for what?" The solution will be simply the primary good, what is good for its own sake, or, in more Platonic terms, the Good itself.

II.

When in *Republic* VI Socrates introduces the Form of the Good as the highest object of knowledge (*megiston mathēma*), he repeatedly insists that "you have often heard this before" (504E7–8, 505A3).

You have many times heard that the Form of the Good is the greatest object of knowledge, and that it is by the use of this that justice and the rest became useful and beneficial. . . . But we do not know this adequately. And if we do not, you know there is no benefit in our knowing all the rest without this, just as there is no benefit in our possessing anything without the good. . . .

⁴ Epistēmēn paradidonai autēn heauthēn at Euthydemus 292D4 recalls "knowledge of knowledge" in the Charmides. Also Euthydemus 292C6-D1 closely parallels Charmides 173D-174B; cf. shoemaking at Charmides 173D9 and Euthydemus 292C8.

Digitized by Arya Samaj Foundation Chennai and eGangotri Furthermore, you know this too, that most people think the good is pleasure [as in the *Protagoras*], whereas the more refined sort think it is wisdom (phronēsis) [as in the Euthydemus and the Meno]. . . And those who hold the latter view are not able to indicate what kind of wisdom, but they end by being forced to say: knowledge of the good. . . [This is ridiculous] if those who complain that we do not know the good go on to speak to us as if we know it. For they say it is knowledge (phronēsis) of good, as if we understood what they say when they pronounce the word 'good' (Republic VI, 505A-C).

I want to suggest that the passages we have been considering here from the premiddle dialogues—from the Laches, Charmides, Lysis, and Euthydemus—provide the literary background for Socrates's claim that "you have heard this many times before." What we have is a reference neither to Socrates's historical practice nor to oral instruction in Plato's school, but a reference wholly within the fictitious art world of the "Socratic" dialogues, a reference to those passages in the earlier dialogues which present the knowledge of good as making all things useful and beneficial. And this is as close as we can reasonably expect Plato to come to an explicit avowal that the passages in question were all designed to be read proleptically, by reference to the doctrine that is finally stated here in Republic VI.

So far we have been tracing the proleptic function of the "Socrates's dream" passage in *Charmides*, where the knowledge of good and evil replaces the reflexive knowledge of knowledge as cause of public and private happiness (173A–175A; cf. 171D ff.). I want to conclude by an analysis of the preceding section of the *Charmides*, where it is argued that knowledge of knowledge, understood as knowing what one knows and what one does not know, is something impossible or useless or both (169D–171C). What is striking here is (a) that this conception of self-knowledge is formulated in terms that refer unmistakably to Socrates's own practice of testing the knowledge claims of his interlocutors, as reported in the *Apology*, and (b) that this conception is rejected here as incoherent, on the grounds that the examiner can only test knowledge claims in a field in which he is himself an expert.

The critique of Socratic self-knowledge (170A–171C) relies upon the principle that each particular art or science is defined and individuated by its subject matter: there is a one-to-one mapping between a specific kind of knowledge and a specific subject matter (*Ion*

⁵ The key word exetasai recurs at 170D5, 172B6-7. Cf. exetasis at Apology 22D6, exetazein at 23C4-5, 38A5, 41B5, etc.

537C-538A, briefly reformulated at *Charmides* 170A10ff., 171A5-9). Hence, if knowledge of knowledge is possible at all (which has been hypothetically assumed, but not established), such a knower will perhaps know that he knows something, but he cannot know what he knows. Still less will he be able to examine others concerning their knowledge. He may able to tell whether or not they know something, but he cannot judge what they know, or whether they know any specific art, unless he knows the first-order subject matter. Only the doctor, who knows the subject matter of medicine, can judge another doctor's credentials. Hence, if temperance is "knowledge only of knowledge and ignorance, it will be unable to judge whether or not a doctor knows his art, if he is pretending or sincerely believes he knows, nor will he be able to judge any others who know any subject whatsoever, except of course for those who share his craft (homotechnos), like any other specialist" (171C).

How good is this argument? Or how good did Plato think it was? First of all, there is some doubt whether the notion of second-order knowledge without first-order content is defensible at all—whether it corresponds to epistemology, perhaps, or to something like philosophy of science. The dialogue is noncommittal on this point (169A & D, 175B6-7); this is not where Plato's interest is focused. But, assuming the possibility of such second-order knowledge, the preceding argument must be evaluated with regard to four different cases:

- 1. X knows that he possesses the knowledge of F (e.g., of medicine).
- 2. X knows that he does not possess the knowledge of F.
- 3. X knows that Y possesses the knowledge of F.
- 4. X knows that Y does not possess the knowledge of F.

The first case, knowledge of one's own first-order knowledge, is not paradoxical, and is implicitly allowed for in the argument at 170B6-7. (See also 172B.) Plato is careful (at 170A2-4) to distinguish this from the second case, which he explicitly represents as incoherent: "it is impossible to know somehow or other what one does not know at all" (175C5). Here we might suppose that Plato is exploiting, or is victim of, an ambiguity of reference. For I really can know of some subject F that I am ignorant of F. Thus, without knowing quantum mechanics I can know enough about quantum mechanics to know that I am ignorant of it. Of course, I must know something about it besides the name, or I could not be sure of my ignorance; I cannot be wholly ignorant of the subject. So the argument is not without a point even in this case. Knowing one's ignorance of subject F implies

Digitized by Arya Samaj Foundation Chennai and eGangotri some first—order knowledge of F after all. But it is above all to cases 3 and 4, to the Socratic enterprise of examining his interlocutors for their knowledge or ignorance, that the argument cogently applies. I cannot test the credentials of an alleged expert in quantum mechanics or computer science unless I know quite a lot about the subject in question.

It seems that for cases 3 and 4 the argument is sound. If it is, or if Plato thought it was, the implications for Socratic self-knowledge are considerable. The first implication is that, if Socrates can test his interlocutors for temperance, he must himself be temperate. More significant is the implication that Socrates cannot successfully test for knowledge in any subject in which he is himself ignorant. So his claim here and elsewhere to have only second-order wisdom, the knowledge of his own ignorance, is not to be taken at face value. Insofar as Socrates undertakes to expose the ignorance of others in a matter where he disavows knowledge, either the elenchus must be fraudulent or the disavowal must be ironical and ultimately insincere. 7 Insofar as the elenchus succeeds in revealing genuine ignorance in the interlocutors, Socrates must himself possess the relevant sort of first-order knowledge; or so the argument of Charmides 170A-171C clearly implies.

What is the knowledge which Socrates possesses and which his interlocutors generally lack? The final section of the dialogue gives us a clue. If he successfully examines his fellow Athenians concerning virtue and the good life, Socrates must himself know something about human excellence and, more generally, about what is good and what is bad. In the Apology, any claim to such knowledge remains hidden behind the ironic mask of ignorance. What the Charmides shows is that this cannot be the whole story. The success of Socrates's elenchus presupposes his own possession of the relevant form of knowledge or expertise.

Surprisingly few commentators have recognized that this argument constitutes a serious critique of the Socratic avowal of ignorance. One recent writer who does recognize this sees Plato as indi-

Ironical surely in the context of the Charmides. I do not venture to say how the Apology is to be interpreted. Space does not permit discussion of Gregory Vlastos's important article, "Socrates' Disavowal of Knowledge," The Philosophical Quarterly, xxxv, 138 (1985): 1-31.

⁶ There is a significant link here with the paradox of inquiry in *Meno* 80D-E: How can one investigate what one does not know at all? Both the problems of priority of definition and the knowledge of one's own ignorance can be seen as special cases of the generalized paradox of inquiry.

cating here the limitations of the Socratic elenchus as a philosophical method.8 I do not believe, however, that the Socrates of the premiddle dialogues is identified with any single method of argument. In these works the figure of Socrates stands midway between the Socrates of the Apology (which is probably as close to the historical Socrates as we are able to come) and the Socrates of the Republic, the spokesman for Plato's mature philosophy. His position is essentially unstable, moving back and forth between these two roles. One function of the premiddle dialogues is to raise the questions concerning virtue, knowledge, and education to which the theory of the middle dialogues will respond. Given the literary form of these works, it must be the task of Socrates both to ask the questions and, eventually, to formulate Plato's response. By putting the critique of Socratic ignorance in Socrates's own mouth, Plato is not so much criticizing the figure of Socrates as endowing him with the full prerogatives of the true (i.e., the Platonic) philosopher. Plato must have seen his own theory as providing the necessary foundation for the moral and intellectual stance of the historical Socrates. In this sense, Platonic metaphysics and epistemology can be thought of as Plato's answer to the question: What kind of knowledge is required for the success of the Socratic elenchus? The critique of Socratic ignorance in the Charmides is designed to make clear that, for Plato, this is a question that calls for an answer.

CHARLES H. KAHN

University of Pennsylvania

⁸ R. McKim, "Socratic Self-Knowledge and 'Knowledge of Knowledge' in Plato's Charmides," in Transactions of the American Philological Association, CXV (1985): 59-77.

THE JOURNAL OF PHILOSOPHY

Digitized by Arya Samaj Foundation Chennai and eGangotri

UNIFYING PLATO*

S there a unifying theme, argument, or even system in Plato's dialogues, or is his literary corpus more a sum of parts than a whole? Are the dialogues like pieces of a single jigsaw puzzle, or like pieces of different puzzles? Does it matter to our understanding of Plato which dialogue we begin with and which we end with? These are difficult and important questions, ones which have puzzled commentators since antiquity. The answers will affect, among other things, one's judgment as to whether Plato has a "philosophy" in the sense that Aristotle or Kant or Husserl may be said to have philosophies.

Charles Kahn's "proleptic" reading of the "premiddle" in relation to middle dialogues is oriented by the order in which he thinks they were composed. He then seeks to coordinate this chronological order with the development and resolution of certain themes. In this way, at least part of the corpus is unified into a coherent research project.

In principle, it is possible that these dialogues are so unified. There are some important difficulties to be overcome, however, if a convincing case is to be made.

Kahn tells us that genuine prolepsis in an earlier dialogue must be visible to the "original reader" who did not have access to the anticipated solution because the requisite dialogue was not yet available. This stipulation makes some controversial assumptions about the ways in which Plato's views were disseminated (I refer to the problems of the oral teaching, and of the sense in which the dialogues were "published"). Perhaps the "original reader" should be thought of as the student embarking for the first time on a reading of Plato's texts. Yet Kahn makes any such construal difficult in also holding that "the meaning of a Platonic text must be something that is in principle accessible to the original reader." Perhaps this statement means, however, that our interpretation of a Platonic text is not dependent upon (though it may be aided by) our understanding of another Platonic text, i.e., that the individual dialogues are the primary wholes for the interpreter. I agree. But then prolepsis (as distinct from aporia) in Kahn's sense has little bearing on the meaning

^{*} Abstract of a paper to be presented in an APA symposium on Plato and the Proleptic Reading of Socratic Dialogues, December 29, 1988, commenting on a paper by Charles Kahn, this JOURNAL, this issue, 541–549.

of a Platonic dialogue. The "original reader" standard, in sum, is

puzzling.

The matter is further complicated by Kahn's statement that "the full meaning" of a premiddle dialogue is visible only when it is taken together with that group of dialogues "as a whole." Given that the dialogues within this group were not composed simultaneously, the meaning of each of these dialogues could not have been "in principle accessible to the original reader" if each must be understood in light of the whole group.

Because the Platonic texts are dialogues in which Plato does not speak directly, it is very difficult in any case to determine that a *Platonic* solution has been offered. In spite of his insistence in other papers on the importance of Platonic anonymity and the lack of "transparency" in the dialogues, Kahn asserts without argument at the end of the present essay that the Socrates of the *Republic* is "the spokesman for Plato's mature philosophy." Even if Socrates were Plato's spokesman here, it would have to be shown that what Socrates says amounts to a doctrine or solution to a problem rather than to the restatement of the problem (albeit at a deeper level). And even if all that were shown, we would also have to demonstrate that the given theory of the later dialogue is intended as a solution to the aporia stated in another, earlier text, and conversely that the aporia of the earlier text calls for just the solution offered in the chronologically later dialogue.

Further, Kahn's example of "meaningful prolepsis," namely the development of the "knowledge of good and bad" theme, fails on his own account when he claims that "only the modern reader" looking backwards from the *Republic* to the premiddle dialogues would understand the prolepsis. I shall argue that the "you have often heard this before" citation from the *Republic* does not provide any internal support for even the retrospective reading, and, more importantly, that the *Charmides* does not present us with an "essentially unstable" Socrates who vascillates between questioning and answering.

I shall also suggest that the dialogues—in part because they are dialogues—are all proleptic, whether or not they are explicitly aporetic, in that they always withhold a final resolution of the given problem. They are self-consciously partial reflections. Their partiality is not completed by other dialogues so much as by the reader's reflection on the whole nature of the matter discussed. In more Platonic terms, individual dialogues point to the whole rather than to the wholeness of a literary corpus.

CHARLES L. GRISWOLD, JR.

Howard University

Digitized by Arya Samaj Foundation Chennai and eGangotri ARGUMENT AND PERCEPTION: THE ROLE OF LITERATURE IN MORAL INQUIRY*

LMOST all contemporary philosophers share a core conception of moral inquiry, which I shall call the intellectualist view. The shared conception holds the most important feature of moral inquiry to be examining arguments: the investigation of the logical and evidential relationships among believed propositions. Which propositions should serve as premises for these arguments is hotly debated. Some endorse coherentist approaches that allow a person to use all her initial beliefs, revising where contradictions or incoherencies are discovered. Others think there is an a priori starting point to be found in the meaning or logic of moral language, or the theory of rational decision. Still others hold our main task is to find a place for values within the world revealed by the natural and social sciences. But the fact is, all of these approaches share the common view that philosophical inquiry into morality is preeminently a matter of investigating the logical and evidential relationships among the relevant propositions, whether these relationships be deductive, inductive, abductive, or what have you.

It is doubtful anyone holds this is all there is to moral inquiry. Inquirers must be allowed to acquire additional beliefs, with more facts being especially welcome. Most philosophers recognize new moral and philosophical beliefs must be permitted as well. But, according to the intellectualist view I wish to identify, the acquisition of such beliefs is a ticklish matter. There is a strong temptation to think the only way moral and philosophical beliefs can be added or altered legitimately is where the addition or alteration is somehow required either by beliefs that are already present or by the acquisition of new empirical information.

This conservative element, combined with the focus on arguments, suggests the intellectualist takes mathematics and science as models for moral inquiry. We are invited to think of the moral theorist as occupied with one or more of the following: sorting through the "data" to locate and eliminate any aberrations (whereas for some the data will be restricted to facts and logical or linguistic intuitions, for others the data will include intuitive moral judgments as well), at-

0022-362X/88/8510/0552\$01.40

© 1988 The Journal of Philosophy, Inc.

^{*} To be presented in APA symposium on Argument and Perception, December 30, 1988. Richard Brandt will comment; see this JOURNAL, this issue, 566-568. I am indebted to a number of people, but I must mention in particular David O'Connor and Philip Quinn. I have also benefitted from Martha Nussbaum's articles in more ways than my scanty references reveal.

Digitized by Arya Samaj Foundation Chennal and eGangotri tempting to construct a theory that adequately explains the data, and deducing theorems or applications to particular cases. Of course, the mathematical/scientific model is not completely adequate, and some intellectualists might quibble with the model. And the model itself is problematic, since there is no settled position regarding the nature and epistemology of either mathematics or science. Nonetheless, it is helpful to regard the intellectualist conception of moral inquiry as based on a mathematical/scientific model. A large number of theorists explicitly cite such a model, and those who do not propose a method of moral inquiry that "looks" scientific. Accordingly, debates about the rationality of these methods focus on establishing similarities or dissimilarities with science or mathematics.

Lest anyone doubt the power of the spell woven round ethics by intellectualism, consider Richard Brandt's position. Brandt defines rational desires as those which would survive cognitive psychotherapy: a "process of confronting desires with relevant information, by repeatedly representing it, in an ideally vivid way, and at the appropriate time" (ibid., p. 113). Although the issues surrounding this notion are complex, my purpose is not to broach them here. I wish merely to use Brandt to illustrate how thoroughly philosophers have fallen under the sway of the intellectualist view. For although it is certainly clear, as Brandt argues at length, that a person's desires can be altered by repeated reflection on relevant information, it is astonishing that anyone could take such alteration of desire to be central, let alone propose ability to survive mere reflection on the facts as a criterion of rational desire. This proposal is particularly shocking given that Brandt himself notes both that certain desires "may resist extinction by inhibition and anything else, since they have been so firmly learned at an early age" and that his definition entails that such desires are rational (ibid.). To see that only a philosopher captivated by the intellectualist conception could find Brandt's definition of rational desire even initially plausible, one need only reflect on the fact that one plausible example of this type of desire is the desire of persons who were sexually abused as children to abuse children in their turn. Only a discipline bewitched by something like the intellectualist view could spawn a proposal that entails such desires are rational.

My goal is to contribute toward showing that the intellectualist conception of moral inquiry is inadequate. My intention is not to argue that intellectualism is completely misguided. The things intel-

¹ A Theory of the Right and the Good (New York: Oxford, 1979).

lectualism focuses on do have a signification the redering the redering the property in quiry. All I wish to suggest is that the picture of moral inquiry the intellectualist provides is not complete, that it will not allow us to understand crucial features of moral inquiry. Do not underestimate the importance of this conclusion: a conception of moral inquiry allowing even a significant role to the elements recognized by the intellectualist can still diverge widely from the intellectualist conception. My strategy for the paper is as follows. I begin with examples illustrating two phenomena: naiveté and the corruption of moral judgment. The examples also suggest that literature, theater, film, music, and art can play a role in both relieving naiveté and causing corruption. I then attempt to explain the epistemological features of these phenomena within the framework supplied by the intellectualist conception of moral inquiry. Although we can go a considerable way working within the intellectualist conception, however, we cannot explain all the features of the phenomena. I close by suggesting that abandoning the argument-based intellectualist approach and adopting a perceptual model allows a more adequate understanding of naiveté, corruption, and the significant role of literature in moral inquiry.

TWO PHENOMENA: NAIVETÉ AND CORRUPTION

Imagine a developer who, being interested solely in the exploitation of land for economic gain, does not see unspoiled natural areas as good at all. Just as this person cannot see any good in wild places, he cannot see good in pastoral areas that have been benignly turned to agriculture. Our developer recognizes the need for agriculture and he knows agriculture is a major industry. But he is not influenced by any sentimentalism about family farms, about there being intrinsic value in this particular form of life, nor does talk of living in harmony with the land strike any chord in him, and he certainly would see no point in trying to run a small, diverse farm relying as little as possible on chemicals. Our person proudly pictures himself as a "bottom line" type, interested in high yields with a minimum amount of labor, fields as large as possible planted with the most valuable crops, and animals in feedlots rather than on pasture where they burn calories uselessly walking around.

Our developer could incorporate his beliefs regarding the moral insignificance of the natural environment into a coherent system of moral beliefs, thereby becoming immune to any efforts to argue him out of his views. For if we think only in terms of human interests, even taking a long view, the issues are complex. And we can suppose the developer has invested a considerable amount of time studying questions concerning land development and agriculture. He could understand the relevant facts, know about the effects of various

farming methods and development schemes, and be acquainted with the arguments offered by environmentalists. But knowing how the land will be changed by, for example, clear-cutting a forest or putting up a shopping mall would not force our developer to alter his stand. This knowledge would generate no internal conflict, since he would not feel the slightest inclination to think, e.g., that a tropical forest is somehow better in itself than rivers clogged with silt from rapidly eroding mounds of soil too poor to sustain significant plant growth. And even if he could be convinced that "organic" farming methods can turn a profit, he would see no point in accepting less than maximum profit so a given plot of land is not a "biological desert" inhabited by a single hybrid species of soya bean. So, the arguments offered by environmentalists could find no foothold within the developer's belief system, since many of these arguments rely on the rejection of the anthropocentric value system that represents his most deeply held moral and philosophical principles, and the remaining arguments can be answered with a dizzying array of empirical data and counter argument. Thus, engaging in moral inquiry as the intellectualist conceives it would not lead the developer to alter his views.

Supposing our developer's thinking does not represent self-interested sophistry, it is likely to seem to one within the grips of the intellectualist model as if his beliefs are unobjectionable. I do not mean the developer's substantive moral views will seem correct. My concern is epistemological, and what I assert is that the intellectualist will not be disposed to see epistemological grounds on which to criticize the developer. His views may be false, but the intellectualist will see the developer as having done all he can by producing arguments for his view and meeting the arguments against it. He will have accomplished what he can working with facts and logic alone.

As a matter of fact, however, we do not yet know enough about the developer to evaluate his moral beliefs. Let us fill out the description: the developer has had no first-hand experience with wild places or things. He has never farmed nor even grown a backyard garden, and he has never visited a traditional family farm, a large "industrial" farm, or an "organic" farm. He has never spent time in a wilderness, nor does he watch nature films. He has read no literature that might convey an appreciation of the value of the natural world or agrarian life style, e.g., the poems of Wordsworth, Sand County Almanac by Aldo Leopold, or Hardy's Wessex novels.

If the intellectualist model has not completely permeated our thinking about moral epistemology, our confidence in the credentials of the developer's beliefs should be shaken by the supposition that his actual and imaginative experience with the things about

which he judges is so severely limited. The developer's arguments may be as clever as we could wish. His mastery of the relevant facts can be ever so impressive. The elegance of his theoretical explanations, moral, philosophical, and empirical, might be a marvel to behold. But all of this intellectual sophistication cannot make up for his experiential naiveté.

The difficulty posed by naive inquirers might not seem serious, for it may not seem that we must stray from the intellectualist view to overcome it. After all, one should not form judgments hastily: a person must (ideally) form judgments on the basis of all the relevant evidence and only after considering the alternative theories as well as what can be said in their behalf. Perhaps we have been too narrow in our thinking about the form in which "arguments," "evidence," and opposing "theories" can be presented, so that we have mistakenly excluded novels, poems, autobiographical accounts, films, and certain kinds of life experiences. But, the intellectualist will think, there is no problem seeing how to correct our oversight. We must simply include what we have left out. We shall continue to think that the most significant thing about moral inquiry is attaining the kind of coherent, highly articulated system of beliefs which emerges from the argumentative process so highly prized by the intellectualist, and simply add that, in the ideal case, a person must expose himself to all the relevant "data," "theories," and "arguments," no matter what their form. Of course, no one can reach this ideal, but even so, we recognize that, if a person falls far enough short, the epistemic credentials of his beliefs will come under suspicion.

Although this approach to the problem of inadequate moral experience has intuitive appeal, it is inadequate. To show why, let me describe an example illustrating a related phenomenon. Imagine a more or less ordinary young man. He tends toward liberal, egalitarian views about appropriate relations between men and women. Wishing to ground his judgments on a firmer foundation, the young man decides to read On the Subjection of Women, and finds himself in substantial agreement with J. S. Mill. Our young man has had little experience with women, however, and realizes he is in a poor position to be making judgments about what sorts of relationships are best. The young man correctly decides more philosophical argument will not significantly improve the epistemological status of his judgments, but makes a very poor choice about how to broaden his imaginative experience—he picks up a romance novel. As a result of reading this novel, the young man comes to think that what a woman really wants is for a man to "sweep her away," that women need to be dominated by men, both emotionally and sexually, and that rough treatment, bordering on rape, is often appropriate and even desirable. If the young man were to stop with this one novel, he would be just as naive as when he began, so he undertakes to broaden his exposure in many directions, reading lots of novels, essays, and poems and watching plays and films. Unfortunately, that first trashy romance has done its work. Accompanied by an appealing glow, images of the seductions in the novel stick in his mind. Throughout his reading, what resonates with his sensibilities are only those things which confirm the perspective he originally picked up in the romance novel. We can plausibly imagine the same thing in his life—the relationships and sexual encounters the young man finds most rewarding are those most closely approximating the paradigm laid down in the novel. The end result is that the young man's moral judgments regarding appropriate relations between men and women are thoroughly fouled up.

I hope this example does not seem far fetched. For although it is oversimplified, what I am trying to describe goes on all the time. People pick up paradigms for relationships from the worst sorts of novels and films, and then tragically try to recreate these paradigms. More to the point, the corrupting influences of such literature and art on a person's judgments can be tenacious, defying all efforts to dislodge them by conveying additional empirical information or confronting the person with philosophical arguments. We should not be too smug about this example, thinking that this kind of corruption is exceptional. For surely we all recognize that there are experiences that would probably lead us radically to alter our system of values. Yet, for the most part, we do not feel that it would be a good idea to go out and have these experiences.

WARRANT AND AN INTELLECTUALIST ACCOUNT OF THE PHENOMENA If my examples have been successful, they illustrate the possibility of naiveté and corruption, as well as the potential of literature² to relieve the former and to engender the latter. Anyone engaged in moral inquiry must be concerned with enjoying a sufficiently broad range of experience and with whether she has had the right kinds of experience. The naive person needs to go out and "live a little" or perhaps to read various novels and poems. But the inquirer must be selective. Experiences are not benign, as facts are, so that we should become familiar with as many as we can. Some have the potential to corrupt a person's moral judgment, lowering the epistemological status of the person's moral beliefs. But whereas my examples illustrate these potential epistemological problems with our moral beliefs, they have not made clear the precise nature of the epistemological deficiencies of naive and corrupted moral beliefs, nor how experiences.

² Henceforth I shall limit myself to literature, but I would want to include films, plays, and even music and art.

rience, particularly that provided by literature, might help one overcome these deficiencies. To make progress in these matters, we must distinguish between two important dimensions along which we make epistemological assessments of belief.3 On the one hand, we are concerned with the rationality of a person's beliefs. The relevant ideal of rationality is of a person doing all he can, as determined from his own point of view, to insure his belief is true. On the other hand, we are concerned with knowledge, and more particularly, with warrant, which is the positive epistemological status that plays the preeminent role in distinguishing mere true belief from knowledge. There is obviously a great deal of debate about what might play this role. But it would hardly be appropriate to set about trying to settle these disputes in the present context. I propose, therefore, that we begin with the following minimal assumption about warrant. As opposed to rationality, which is an internalist conception in that it holds a person to standards that are in some sense his own, warrant has a strong externalist component. Thus, I assume that having warranted beliefs involves satisfying an "objective," external standard that the believer may not accept or even be aware of. In seeking to understand the epistemological deficiencies of naive and corrupted inquirers and the epistemological contribution of literature to moral inquiry, one can focus on either rationality or warrant, but I shall limit the discussion to the latter 4

One view about how a person's moral beliefs might be warranted is suggested by the intellectualist conception of moral inquiry. Given that this conception construes moral inquiry as primarily concerned with arguments, i.e., with examining the logical and evidential relations between believed propositions, one enamored of this conception will be inclined to think of the justification of moral beliefs primarily in terms of the transfer of warrant. For whereas examining the relations between believed propositions and revising beliefs accordingly is widely held to be capable of increasing the epistemological status of the beliefs involved, few hold that this process alone could bring beliefs to be warranted no matter what their initial epistemological status.⁵ Accordingly, intellectualists who think moral theories must be constructed on the basis of facts and logic will think our moral beliefs will be warranted just in case we can construct arguments that are strong enough to transfer sufficient warrant from the relevant facts and logical truths.

⁴ I have cut a treatment of the phenomena in terms of rationality to meet constraints on space.

These distinctions rely heavily on the work of William Alston, particularly "Concepts of Epistemic Justification," The Monist, LXVIII, 1 (1985): 57-89; Richard Foley, The Theory of Epistemic Rationality (Cambridge: Harvard, 1987); and Alvin Plantinga, Warrant, in progress.

If we think of our moral beliefs being warranted as a result of such a transfer of epistemological status, as is natural for the intellectualist, how should we account for naiveté, corruption, and the role of literature in moral inquiry? Discerning the general shape of the account is easy. The account will assume experience is epistemologically interesting primarily as a source of information. Ideally, a person should be acquainted with all the information or evidence before making a judgment, but it is never possible for a person to meet this ideal. We must settle for less. One approach to determining how much information is enough is to say that a certain body of evidence warrants a given conclusion only if it also warrants believing that the total evidence would support the conclusion. According to this understanding, even though the information provided by a naive person's experience might support his moral judgments, that experience is too limited to warrant claiming that broader experience would continue to support them. Given the externalist element of warrant, the problem is not that the beliefs are based on an amount of experience the believer considers to be insufficient, but that the experience is objectively insufficient, whether the inquirer realizes this or not. It might seem that we do not need to think of moral experience in any more complicated way to account for corruption. Information or evidence can be misleading, since it is possible for true propositions to support a false conclusion, and even a conclusion that would be disconfirmed by the total evidence. The intellectualist will want to say that what I have called corruption is nothing more mysterious than misleading evidence.

The intellectualist takes the problems of naiveté and corruption to revolve around information. So apparently the solution to these problems is more information. One might suppose the relevant information can be conveyed by such things as reports of observations or experimental results, but the example of the developer suggests otherwise. Perhaps some information must be acquired either through direct experience or through imagination guided by literature. It is easy to see how literature could play an important, though not essential, role in relieving naiveté. In the first place, a novel can convey new information to a person. For example, Charles Dickens's novels convey a great deal of information about the conditions of the poor in 19th century England. Although the intellectualist will see the most clear cut contributions of literature as involving information, there is no need for him to ignore other potential epistemological contributions. Thus, secondly, the intellectualist need not deny

⁵ Pure coherentists would be sympathetic to this view. Cf. Lawrence Bonjour, *The Structure of Empirical Knowledge* (Cambridge: Harvard, 1985). Note, however, that even on such a view there is a place for the transfer of warrant.

Digitized by Arya Samaj Foundation Chennai and eGangotri, that novels can function in the same way as the more familiar, abbreviated examples philosophers usually employ to elicit linguistic, or even moral, intuitions. He might even think the intuitions a person has about a novel are more reliable, either because the case is more true to life and described in more detail, or because the proper object of moral judgment is a larger slice of a life than philosophical examples ordinarily portray.

Finally, Martha Nussbaum has suggested that certain philosophical views cannot be presented accurately in the form of a philosophical treatise. 6 She argues that the Greek tragedies should be taken to represent such a conception of the good life: a conception fundamentally different from Plato's which he sought to defeat. If Nussbaum is right about the tragedies, other literary works might also represent unique moral conceptions. As striking as this view is as a position regarding the connection between philosophical content and form of presentation, however, it does not require any new epistemological thinking. For as we have noted, ideally a person should consider all the relevant alternatives to his position before making a final judgment. One might not have thought this would involve reading Aeschylus and Sophocles in addition to Plato and Aristotle, but, if Nussbaum is right, it does.

It might then appear that we can account for the phenomena of naiveté and corruption and the role of literature in moral inquiry within the intellectualist conception of moral methodology. This appearance is inaccurate, however. For one thing, if the transition from being morally naive to being a mature moral judge were solely a matter of acquiring information, we could not account for corruption. The analogy with misleading evidence is rough, failing to capture a significant fact-corruption can be permanent, but there is always an antidote to misleading evidence, viz., more information.

More importantly, the intellectualist account is just too simple. A novel can provide a person with new information. And I would not deny that novels can both function as do the more simple examples common in philosophical literature and present alternative philosophical conceptions. But these functions do not exhaust the roles literature can play in moral inquiry. If they did, and we do not go along with Nussbaum in supposing some moral conceptions cannot be presented in a standard philosophical style, then the role of literature in moral inquiry would be purely heuristic. For consideration of literary works would not do anything that could not be done, perhaps more cumbersomely, within the confines of traditional philosophy. The interesting case of naiveté, however, is one where the person has

⁶ The Fragility of Goodness (New York: Cambridge, 1986), pp. 12-18.

a coherent and comprehensive system of moral, philosophical and empirical beliefs supporting his moral judgments. The person has no belief one could use to argue him out of his other beliefs. In order to disrupt such a person's settled convictions and force him out of his naiveté, the person must be brought to have additional experiences. The kind of experience I have in mind can be rather difficult to live through, but fortunately they come infrequently in actual life. I maintain that literature provides a remedy for both features. If I am right, and exposure to literature can bring a person who has led a rather dull, morally colorless life to a high level of moral sophistication, it is a gross distortion to think of literature as serving simply a heuristic function. For literature would be doing something philosophical argument could not do.

LITERATURE AND PERCEPTION

How then should we understand the experience of reading a novel? What other role could literature play in the warrant of a person's moral beliefs? A perceptual model can help answer these questions. The model also enables us to deepen our understanding of naiveté and corruption. Thinking of our moral beliefs as produced by a faculty roughly analogous to perception is nothing new, but there is a tendency to think in terms of the wrong sort of perception, i.e., of perceiving simple qualities. But I am not thinking of things like seeing that a book is red or tasting the saltiness of a pretzel. Rather, I have in mind our capacity to make fine perceptual discriminations, where typically a considerable amount of experience, and sometimes training, is needed to develop the capacity. Think then of the ability of connoisseurs to make fine discriminations between wines or the ability of a biologist to "see" through a microscope.

It will be helpful to have before us an example of the kind of perception I wish to use to understand moral inquiry. After marrying an equestrienne, I have spent a great deal of time with equines, which has provided me recent first-hand experience with perceptual training. Here is a specific example. Horses have three natural gaits: walking, trotting, and cantering. Cantering is a bounding three-beat gait. The horse first moves forward a back leg, then the other back leg along with the diagonal front leg, then the remaining front leg, followed by a moment of suspension. A horse can canter in two different ways: depending upon which front leg remains forward, the horse is said to be cantering on the left or the right lead. Now, the interesting thing about cantering is that even after you get the basic idea, it is not easy to "see" a cantering horse's lead. Along thunders the beast, legs going every which way, and it is just very hard to sort it all out. I can remember my wife repeatedly explaining the idea of a lead, and being exasperated because I could not tell a horse's lead. But I understood the concept of a lead perfectly well, and it was not that I lacked any information. The problem was that my eye could not sort out all those legs; I was unable to see what I had to if I was to apply the concept. After some experience watching horses I started to get the knack. Still more experience and I could immediately "see" a horse's lead in just about the same way I could all along see whether a horse was walking or trotting.

In a sophisticated style of riding, such as dressage, there is much more to see than what lead a horse is on. Not all of the things one can learn to discriminate are as purely "objective" as a horse's lead. One learns to discriminate and evaluate together: to see, e.g., the quality of a horse's gaits or whether the horse is correctly "on the bit." It takes a great deal of time and experience to develop the relevant perceptual capacities, and perhaps no amount of experience or even training could bring the perceptual capacities of some up to the level of a German *Bereiter*. But it is not at all difficult to notice the development of one's own discriminative faculties.

There is a natural way to think about my story. I have a faculty of visual perception, but this faculty did not come to me fully developed, able to make all the discriminations it might. Perceptual experience is required to be able to make these discriminations. Although it is unclear exactly how such experience functions, it clearly does not simply supply information. The experience does not serve merely as additional input to the perceptual faculty that processes this input and yields a belief as output. While also serving as such an input, the experience must somehow alter the nature of the perceptual faculty so that it functions differently, yielding output for a given input that is different from the output that input previously evoked. Of course, it is not just that the perceptual faculty functions differently. As a rule, with experience it functions better, producing beliefs with more warrant, or producing warranted beliefs where previously it produced beliefs without warrant.

The perceptual model provides a natural way to think about naiveté and moral experience. The problem with a naive moral judge is not that there is any information he lacks, or that the system of theories and concepts he has built up is somehow inconsistent, or that he has failed to consider some alternative theory. The problem is that a significant number of the beliefs he is using to develop his moral theory are the products of a faculty of moral judgment that has not been used on a sufficient number of different types of situations or objects, or perhaps not been used on the right kinds of situations or objects, for it to develop to the point where it is competent to make the sorts of discriminations it is being used to make. The result is that the person's moral judgments are not warranted.

One significant function of literature is to supply the kind of experience needed to develop a person's faculty of moral judgment. It is easy to see several ways in which novels can serve this function. First, the kinds of complex situations which require careful moral judgment do not arise all that often in a person's life. So actual experience is unlikely to provide the sheer number of cases it is likely to take to develop fully a person's faculty of moral judgment. Literature is ideally suited for supplying additional cases on which a person can exercise her moral faculties. Second, when we encounter situations requiring moral discernment in real life, we are often so emotionally involved that it is difficult to determine whether our moral faculty is operating freely or is clouded by intense feeling. Literature supplies cases remote enough to minimize the effects of personal bias and excessive emotion, while still sufficiently realistic to engage our emotions to the degree necessary for the proper functioning of our moral judgment. Third, while the situations described in novels are far more complex than the kinds of examples common in the philosophical literature, they are far less complex than real situations. Situations are interpreted for us to a certain degree in a novel. The author is able to provide some guidance as we attempt to understand the situation he describes. This not only provides an example that might help a person to see how to sort out the salient features of a real life situation for himself, but it provides a manageable task on which to exercise a less than completely developed moral faculty.

I would not underestimate the importance of novels as simply offering us the opportunity to practice thinking about difficult and interesting situations and complex personalities and providing us with examples of how to discriminate salient features of such situations and characters. For it is possible to improve our ability to make fine perceptual discriminations simply by practicing, and of course a little guidance can be helpful. I see no reason to expect our faculty for making moral judgments to be any different. Nonetheless, I recognize that I have taken only the smallest step toward understanding how literature can serve to enhance our ability to make moral judgments. Still, I would like to think that I have taken an important step. For we can make no progress in understanding either the epistemology of our moral beliefs or the role of literature in moral inquiry until we stop trying to force things into the mold supplied by the intellectualist. And I am confident the perceptual model is true to this extent: we have a faculty for making moral judgments, but this faculty does not come to us in its final form. It must be developed. Therefore, an important part of moral inquiry should be concerned with the development of this faculty. This conDigitized by Arya Samaj Foundation Chennai and eGangotrice ception of moral inquiry certainly is not without precedent. Recall Aristotle's claims that the young are unfit to hear lectures on ethics, in part because of their lack of experience, and that only those who have had the proper sort of upbringing will have starting points for moral theorizing. These remarks suggest the view that there are experiences a person must have if he is to be fit to work out a moral theory. And we might go on to say that there are other experiences, some of them characteristically enjoyed by persons who have been poorly raised, that make a person unfit for moral inquiry.

I am also confident that by thinking about perception, about how our perceptual capacities are developed, and about how perception serves to warrant beliefs, we will get some guidance in answering the related questions about moral epistemology. In particular, we will begin to see that experience, and literature, and perhaps even more standard philosophical argumentation and theorizing can function in two importantly different ways. They can provide inputs to our faculty of moral judgment, and they can systematize and extend the outputs of this faculty by argument and the construction of explanations and theories. But experience, literature, and even philosophy can also serve to alter the functioning of our faculty of moral judgment, for better or for worse, and it is this capacity that philosophers have neglected in thinking about moral methodology and moral epistemology.

As a sort of confirmation for the perceptual model, we might note that it provides a natural way for us to make sense of corruption. In other areas where we recognize a faculty for making fine perceptual discriminations which has been developed through extensive experience, we recognize that the faculty can lose its high degree of attunement. Thus, for example, an expert wine taster can become jaded or, through exposure to inferior wines, lose his faculty for making fine discriminations. This kind of corruption was nicely described by Mill, as it can affect a person's ability to enjoy refined pleasures.

Capacity for the nobler feelings is in most natures a tender plant, easily killed, not only by hostile influences, but by mere want of sustenance; and in the majority of young persons it speedily dies away if the occupations to which their position in life has devoted them, and the society into which it has thrown them, are not favorable to keeping that higher capacity in exercise.⁷

Some might think that literature does not have the potential to damage or incapacitate a person's faculty of moral judgment, and that we therefore cannot understand moral judgment by using a

⁷ Mary Warnock, ed., *Utilitarianism and Other Essays* (New York: New American Library, 1974), p. 261.

perceptual model. My inclination is to deny that literature lacks the potential to corrupt, and I hope the example of the young man I presented above was sufficiently plausible to justify my inclination. But, even if literature does not have the power to corrupt, it constitutes no serious challenge to my position. For not every sort of experience having the potential to enhance or develop a perceptual faculty must also have the potential to corrupt it. And life experiences can have the invasive, corrosive power or subtle, seductive power it takes to corrupt. Think, for example, of Sue Bridehead's reaction to the death of her children in Thomas Hardy's Jude the Obscure or of the concerns the Amish have about their children attending "English" schools beyond the eighth grade. The proper way of understanding this kind of corruption is in terms of the disruption or incapacitation of a quasi-perceptual faculty of moral judgment, not in terms of misleading evidence. The corrupting experience alters the functioning of the faculty, so the evaluative judgments it produces are no longer warranted.

CONCLUSION

As I have said, I do not think the intellectualist conception of moral inquiry is adequate. We cannot understand moral inquiry as simply a matter of constructing arguments on the basis of some portion of our antecedent beliefs, whether this portion contains only facts and logic, or it contains intuitive moral beliefs as well. Still, I have tried to see how far the intellectualist can take us toward understanding some basic phenomena in the realm of moral inquiry. I have given the intellectualist conception its best shot, but in the end it proved to be inadequate. In my view, we can go farther in our efforts to understand moral inquiry if we adopt a perceptual model rather than the argument-based model proffered by the intellectualist. If we take this model seriously, philosophical inquiry into morality will take on a rather different appearance. Although there will obviously be a place for argument, explanation, and theory construction, there must also be a prominent place for literature, music, and art and for philosophical reflections on these. To the dismay of many philosophers, there will be a greater role for intuition, but perhaps this will be tempered by more thought being given to the development of intuition.

MICHAEL R. DEPAUL

University of Notre Dame

Digitized by Arya Samaj Foundation Chennal and eGangotri MORAL THEORY AND MORAL EDUCATION*

ICHAEL R. DEPAUL affirms that the dominant view of "moral inquiry" has science as its model, being an investigation of the logical and evidential relations among believed propositions. This, he thinks, is at best incomplete, and proposes that a "perceptual model" be adopted instead.

I do not know if he means to criticize traditions of metaethical inquiry. Here it is true that attention is paid to what value statements mean or express, whether moral statements are sensibly viewed as self-evident like arithmetic, whether, in view of psychology and anthropology, it is reasonable to think that moral beliefs will converge if factual disagreements are eliminated. I do not see that DePaul intends to criticize this.

In saying that the usual method of "moral inquiry" is defective, I think he must refer to the reasoning philosophers recommend as a way to reach moral truth as nearly as possible. But here I do not recognize his description of what contemporary philosophers do. I do not see them as recommending modeling ethical thinking after science. I should not say 'all', for there are some exceptions, perhaps David Gauthier and Alan Gewirth, among others. But I think mainstream moral philosophers give an important role to informed preference or attitudinal response in their conception of how one should decide normative issues. For instance, R. M. Hare and William Frankena think the crucial question is what you can want (or prescribe) universally. This is an appeal to moral reasoners to inform, and inspect, their desires/preferences. Obviously, the same for emotivists like C. L. Stevenson and J. J. C. Smart. John Rawls does not say that the idea of choice in the "original position," when adjusted for conflicts with "considered judgments," is supported by evidence; he appeals to the reader's favoring choices made in the original position, and, when the question arises which principle is to be amended in case of conflict, the answer seems to be preference, or "our moral sensibility." Certainly the same for me: I have suggested that we cannot regard as intrinsically good anything we do not want

0022-362X/88/8510/0566\$00.50

© 1988 The Journal of Philosophy, Inc.

^{*} Abstract of a paper to be presented in an APA symposium on Argument and Perception, December 30, 1988, commenting on a paper by Michael R. DePaul, this JOURNAL, this issue, 552–565.

after we have reflected carefully on our desire, and that what is morally right is what would be permitted by the moral code for our society which we prefer (after careful examination of our desires). And is this not what Kant proposed—that an act is right only if you can will its maxim to be universal law? Much the same, I think, for Derek Parfit and James Griffin.

So what does Depaul's objection come to? I think he is saying that philosophy is too arid and abstract to do the job of moral education, that all its metaethically supported methods for deciding moral issues are unlikely to do the job of moral education in ordinary life.

Take his developer who will not decide to protect the environment on account of a philosophical argument. True, although if he followed the direction of philosophers he might come out differently. I agree with Depaul that where he goes wrong is in the lack of experience which would affect his preferences; he needs to go camping and see what others find valuable in virgin forests. Philosophers at least since Hutcheson have thought a person should bring information and imagination to bear on moral/value responses. He has failed to conform with their recommendations. Then there is the young man who reads a romantic novel and gets the idea that women like to be dominated by men and subjected to rough treatment. His trouble is straight informational; he does not read the newspapers or any women's magazines! No doubt he is fortified in this conception because it feeds his self-interest, but what he needs is an accurate view of what women want, and what it is like to be a woman who is treated in the way he approves. Thus far, it would seem that the trouble with the naive persons he describes is that they have not had enough experiences, and are not able to imagine adequately. And it is not enough for philosophers to recommend that they have more, since they have to decide how much is enough. I agree: I see no way to solve this problem.

But he also thinks philosophers should not recommend all types of relevant experience, for some can be "corrupting." If a person tries a promiscuous life, he is apt to be gripped by it. True. But such a person needs to set alongside this experience some insight into interpersonal relations involving the mind: community of interests, discussion with an equal, joint reflections about life. He has not had too much experience; he has had too little.

For such persons, novels may do more good than dry philosophy. Novels can give vivid insight into what it is like to be another person—say, a woman—and by creating a mood they can make one's emo-

tions more responsive to the plight of others. There is education going on here all right, although we can be spared talk of a "moral faculty." One's imagination and capacity for sympathy can improve, as most philosophers have thought they should. Philosophy has something to offer, however: it can tell us to universalize our judgments; and, if one is trained to do this, one may become sensitive to relevant differences between cases. (Maybe this is what Depaul is asking for in what he calls a "perceptual" model of moral inquiry.) Doubtlessly, not everyone can be "cured" by study of philosophy. But contemporary moral philosophy has something to offer.

R. B. BRANDT

University of Michigan/Ann Arbor

Digitized by Arya Samaj Foundation Chennai and eGangotri NOTES AND NEWS

The London School of Economics and Political Science would like to announce the 1989 Lakatos Award in Philosophy of Science for an outstanding contribution to the philosophy of science in the form of a book published in English during the last six years. The Award is endowed by the Latsis Foundation in memory of Imre Lakatos. Candidates shall be nominated by at least three people of recognized professional standing. Nominators shall give their grounds for the nomination and indicate the candidate's age, since a preference may be given to a younger scholar. Three copies of the book and additional communication may be directed, by April 15, 1989, to the Secretary, London School of Econ., Houghton St., London WC2A 2AE, England.

The Australasian Association for Logic will dedicate its 1989 conference to Arthur Prior, to be held at Prior's alma mater, the University of Canterbury/Christchurch, New Zealand, in August 1989. Papers are invited on Prior and his work, particularly in the areas of tense logic, medieval logic, modal logic, and intensionality. Further information may be obtained by writing Jack Copeland, Phil. Dept., Univ. of Canterbury, Christchurch, New Zealand.

The Scots Philosophical Club is pleased to announce a new series of monographs under the general title *Macmillan Studies in Contemporary Philosophy*, continuing in the tradition of the *Scots Philosophical Monograph Series*. Original manuscripts are invited in all areas of contemporary philosophy, including historical studies, provided they make a contribution to current debate in the subject. Further information may be obtained by writing Andrew Brennan, Phil. Dept., Univ. of Stirling, Stirling FK9 4LA, Scotland.

The Society for Lesbian and Gay Philosophy, founded in April to promote philosophical inquiry into lesbian and gay issues, announces a call for papers for the Spring 1989 Central Division APA meeting. The Gay and Lesbian Caucus was founded at the same time to give a distinct voice to gay and lesbian practical concerns within the Association, such as discrimination in hiring, promotion, tenure, or salary. Profs. Claudia Card and John K. Pugh are co-chairs of both organizations for 1988/9. The Caucus shall receive discrimination complaints and shall issue newsletters concerning activities of the Society and Caucus. The mailing lists of these organizations shall remain confidential. Papers and inquiries may be directed to Claudia Card, Phil. Dept., Univ. of Wisconsin, Madison, WI 53706, or John Pugh, Phil. Dept., John Carroll Univ., University Heights, OH 44118.

The Bertrand Russell Society is pleased to announce its 1989 Doctoral Grant Award, open to a currently enrolled doctoral candidate in any field whose proposed dissertation best gives promise of dealing in a significant way with the thought, life, or times of Russell. Further information may be obtained by writing Hugh S. Moorhead, Phil. Dept., Northeastern Illinois Univ., Chicago, IL 60625. The Bertrand Russell Society also announces a call for papers to be presented at its meeting with the Eastern Division of the American Philo-

Digitized by Arya Samaj Foundation Chennai and eGangotri sophical Association in December 1989. Papers may be on any aspect of Russell's philosophy, and shall be submitted in triplicate in format for blind review, along with an abstract of 150 words, by April 1, 1989, to David E. Johnson, Philosopher's Committee, Bertrand Russell Society, Sampson Hall, U.S. Naval Academy, Annapolis, MD 21402-5044.

The Society for Women in Philosophy is pleased to announce that its 1988 Honorary Woman of the Year is Leigh S. Cauman, Columbia University, former Managing Editor and now a Trustee of the *Journal of Philosophy*. Elizabeth Beardsley was the first honoree (1984), followed by Marjorie Greene (1985), Mary Mothersill (1986), and Elizabeth Flower (1987).

Brown University is pleased to announce the establishment of a Center for the Study of Race and Ethnicity in America. The purpose of the Center is to encourage and support multidisciplinary research and teaching in areas relating to race and ethnicity. The Center hopes to serve as a resource center and as a clearing house for information about scholarship in comparative, theoretical, and ethical studies of race, gender, and class. The Center would be pleased if interested philosophers would write to them, % John Ladd, Phil. Dept., Brown Univ., Providence, RI 02192.

The Heyman Center for the Humanities at Columbia University is pleased to announce a conference on the topic "Mental Representation, Realism, and Research," on December 2–4. Participants include Akeel Bilgrami, Ned Block, Patricia Smith Churchland, Arthur Collins, Arthur C. Danto, Daniel C. Dennett, Jerry A. Fodor, Stevan Harnad, James Higginbotham, Julian Hochberg, Norbert Hornstein, Hide Ishiguro, Jerrold Katz, Eric Kandel, Dan Lloyd, and Herbert Terrace. Additional information may be obtained by contacting Mark Rollins, Phil. Dept., Box 1073, Washington Univ., St. Louis, MO 63130.

The Program in Ethics and the Professions of Harvard University invites applications for Fellowships in Professional Ethics. At least four fellowships will be awarded in 1989–90 to outstanding teachers and scholars who wish to address ethical issues in schools of business, government, law, and medicine. The application deadline is January 13, 1989. Further information may be obtained by writing Valerie Abrahamsen, Program in Ethics and the Professions, Harvard Univ., 79 J.F.K. St., Cambridge, MA 02138.

Santa Clara University will hold its eleventh annual Philosophy Conference on April 7–8, 1989, on the topic "Reason and Moral Judgment," featuring speakers Bernard Williams and Robert Audi. Papers are invited on the issue of rationality and the status of moral justification, and they may be either systematic or historical in content. Papers may be submitted for the conference, the journal *Logos*, or both, in format suitable for blind review, by January 9, 1989. Further information may be obtained by writing William J. Prior, Phil. Dept., Santa Clara Univ., Santa Clara, CA 95053.

AMERICAN APPARE SOUPHICAL ASSOCIATION EASTERN DIVISION EIGHTY-FIFTH ANNUAL MEETING

SHERATON WASHINGTON HOTEL, WASHINGTON, DC DECEMBER 27–30, 1988

PROGRAM

TUESDAY, DECEMBER 27

1:00 p.m.-10:00 p.m. registration, Registration Desk, Lobby Level 1:00 p.m. executive committee meeting, Congressional

Group Meetings (see Appendix for Programs)

 $8:00\,$ p.m. North american division of the schopenhauer society, Richmond

8:00 P.M. CONFERENCE OF PHILOSOPHICAL SOCIETIES, Arlington

8:00 P.M. SOCIETY OF CHRISTIAN PHILOSOPHERS, Dover

8:00 P.M. SOCIETY FOR CLASSICAL REALISM, Baltimore

8:00 P.M. LENSGRINDERS' ASSOCIATION, Marshall

WEDNESDAY, DECEMBER 28

9:00 A.M.-7:00 P.M. REGISTRATION, Registration Desk, Lobby Level 9:00 A.M.-5:00 P.M. BOOK DISPLAYS, Exhibit Hall A

CHILD CARE arrangements may be made on an individual basis with a commercial firm. See the September or November issue of the APA *Proceedings* for further details.

PLACEMENT INFORMATION

For interviewers, 8:00 A.M.-7:00 P.M., Atrium #2 For candidates, 9:00 A.M.-7:00 P.M., Exhibit Hall C

MORNING (9:00-11:00)

Group Meetings (see Appendix for Programs)

9:00 INTERNATIONAL ST. THOMAS SOCIETY, Arlington

9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, Alexandria

9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY AND SOCIETY FOR

THE STUDY OF PROCESS PHILOSOPHY, Dover

9:00 NORTH AMERICAN NIETZSCHE SOCIETY, Wilmington

9:00 WORLD PHENOMENOLOGY INSTITUTE, Baltimore

Digitized by Arya Samai Foundation Chennai and eGangotri Wednesday, December 28, Morning

9:00 international philosophers for the prevention of nuclear omnicide, *Vermont*

9:00 association for informal logic and critical thinking Woodley

9:00 SOCIETY FOR CLASSICAL REALISM, Annapolis

LATE MORNING (11:15-1:15)

11:15 AMERICAN SOCIETY FOR VALUE INQUIRY, Richmond

11:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF GENOCIDE AND THE HOLOCAUST AND THE APA COMMITTEE ON BLACKS IN PHILOSOPHY, Arlington

11:15 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, Colorado

11:15 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY, Warren

11:15 SARTRE CIRCLE, Calvert

11:15 JASPERS SOCIETY OF NORTH AMERICA, Annapolis

AFTERNOON

2:00-5:00 Session I

A. Symposium: History of Philosophy as a Discipline

North Cotillion Ballroom

Chair: Thomas Flynn

Speakers: Alan Donagan and Michael Frede

Commentator: Kenneth Schmitz

B. Symposium: The Philosophical Significance of Experimentation, Wilmington
Chair: Robert Ackermann

Speakers: Ian Hacking and Patrick A. Heelan

Commentator: Peter Galison

C. Invited Papers: Dover

 The Social and Political Philosophy of Thomas Hobbes Chair: David Gauthier Speaker: Jean Hampton Commentator: Joshua Cohen

2. International Justice

Chair: Kwasi Wiredu

Speaker: Bernard Boxhill

Commentator: Gerald Doppelt

D. Colloquium: Heidegger/Critical Theory Debate, Richmond

1. Chair: Babette Babich

Richard M. Capobianco, "Heidegger and the Critique of the Understanding of Evil as *Privatio Boni*"

Commentator: Lawrence Hatab

2. Chair: Hans Seigfried

Linda Alcoff, "Hermeneutics, Coherence, and Truth"

Commentator: Georgia Warnke

3. Chair: Francis Ambrosio

Michael Kelly, "The Gadamer/Habermas Debate Revisited:

The Question of Ethics"

Commentator: Lorenzo Simpson

E. Colloquium: Epistemology I, Baltimore

1. Chair: Monica Holland

Stephen C. Hetherington, "Epistemology's Psychological Turn"

Commentator: Hilary Kornblith

2. Chair: Robert Shope

Michael D. Roth, "The Wall and the Shield"

Commentator: Gary Gleb

3. Chair: David Coder

Richard Fumerton, "Foundationalism and Regress

Arguments"

Commentator: Richard Feldman

F. Colloquium: Ethics, Marshall

1. Chair: Richard Henson

Ann M. Wiles, "Harman and Others on Moral Relativism" Commentator: Gilbert Harman

2. Chair: Judith DeCew

Margaret Walker, "Autonomy or Integrity: A Reply to Slote"

Commentator: Michael Slote

3. Chair: Jesse Mann

Richard Double, "What Is Wrong with Hierarchical

Compatibilism"

Commentator: John Marshall

G. Special Invited Workshop: Charles S. Peirce

South Cotillion Ballroom

Chair: Hilary Putnam

Speakers: Jaakko Hintikka, Vincent Potter, and Richard S. Robin

Digitized by Arya Samai Foundation Chennai and eGangotri Wednesday, December 28, Afternoon, Session 1

H. Invited Papers: Annapolis

1. The Legal and Moral Status of Surrogate Motherhood Chair: Alicia Roque Speaker: Mary Gibson Commentator: Iris M. Young

2. The Uses and Abuses of French Discourse Theory for Feminist Politics

Chair: Dorothy Leland Speaker: Nancy Fraser

Commentator: Arlene B. Dallery

I. Special Symposium Arranged by the APA Committee on International Cooperation, *Alexandria*

Topic: Philosophy in India Today

Chair: Karl H. Potter

Panel: Anindita Balslev, Arindam Chakravarty, D. P.

Chattopadhyaya, Eliot Deutsch, Ashok Gangadean, Bimal Krishna Matilal, Jitendranath Mohanty, and S. S. Rama Rao Pappu

Group Meetings (see Appendix for Programs)

2:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, Arlington

2:00 CAUCUS OF DEPARTMENT CHAIRS, Vermont

2:00 american association for the philosophic study of society, Woodley

LATE AFTERNOON (5:15–7:15)

5:15 society for the philosophical study of marxism, Richmond

5:15 SOCIETY FOR WOMEN IN PHILOSOPHY, Alexandria

5:15 WORLD PHENOMENOLOGY INSTITUTE, Baltimore

5:15 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, Colorado

5:15 AYN RAND SOCIETY, Idaho

5:15 CHARLES S. PEIRCE SOCIETY, Kansas

5:15 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY, Vermont

5:15 INTERNATIONAL BERKELEY SOCIETY, Calvert

5:15 PERSONALISTIC DISCUSSION GROUP, Rockville

5:15 BERTRAND RUSSELL SOCIETY, Holmes

5:15 G. E. MOORE SOCIETY, Annapolis

5:15 ASSOCIATION FOR THE PHILOSOPHY OF THE UNCONSCIOUS

5:15 AMERICAN ASSOCIATION OF PHILOSOPHY TEACHERS, Arlington

APA PROGRAM

Digitized by Arya Samaj Foundation Chennai and eGangotri

EVENING (7:30-10:30)

7:30 ASSOCIATION FOR THE PHILOSOPHY OF EDUCATION, Richmond

7:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, Arlington

7:30 NORTH AMERICAN DIVISION OF THE SCHOPENHAUER SOCIETY

Alexandria

7:30 SANTAYANA SOCIETY, Baltimore

7:30 RADICAL PHILOSOPHER'S ASSOCIATION, Colorado

7:30 LEIBNIZ SOCIETY OF AMERICA, Idaho

7:30 GANDHI-KING SOCIETY, Calvert

7:30 SOCIETY FOR THE STUDY OF MEDIEVAL AND RENAISSANCE PHILOSOPHY, *Annapolis*

7:30 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, Holmes

7:30 INTERNATIONAL SOCIETY OF NEOPLATONIC STUDIES, Marshall

8:00 RECEPTION, Sheraton Ballroom

THURSDAY, DECEMBER 29

9:00 A.M.-5:00 P.M. REGISTRATION, Registration Desk, Lobby Level 9:00 A.M.-5:00 P.M. BOOK DISPLAYS, Exhibit Hall A

PLACEMENT INFORMATION

For interviewers, 8:00 A.M.-5:00 P.M., Atrium #2 For candidates, 9:00 A.M.-5:00 P.M., Exhibit Hall C

MORNING

9:00-11:00 Session II

A. Symposium: Postmodernism and Rationality

North Cotillion Ballroom

Chair: John Caputo

Speaker: Rodolphe Gasché Commentator: Charles Scott

B. Invited Papers: Dover

1. Truth in Aquinas and Ockham

Chair: James Ross

Speakers: Marilyn McCord Adams and John Wippel

C. Colloquium: Husserl, Arlington

Chair: Richard Cobb-Stevens
 Mark P. Drost, "The Primacy of Perception in Husserl's
 Theory"

Commentator: Robert Sokolowski

2. Chair: Izchak Miller

George Heffernan, "Speech and Reason in Phenomenological Logic"

Commentator: Robert Tragesser

Thursday, December 29, Morning, Session II

D. Colloquium: Aristotle's Ethics, Wilmington

Chair: John Cooper
 Howard Curzer, "Aristotle's Treatment of Megalopsychia"
 Commentator: Robin Smith

2. Chair: Mark L. McPherran

Paul Schollmeier, "An Aristotelian Origin for Good Friendship"

Commentator: Douglas Rasmussen

E. Colloquium: Plato and Platonism, Baltimore

1. Chair: Stephen Strange
John E. Peterman, "The Eleusinian Nature of Arete in
Plato's Meno"

Commentator: Nicholas Smith

Chair: Arthur Madigan
 Brian Leftow, "A Platonist Cosmological Argument"
 Commentator: Allan Silverman

F. Colloquium: The Family, Annapolis

Chair: Lawrence Stepelevich
 Heidi M. Ravven, "Hegel and the Politics of Family Life"
 Commentator: Michael Hardimon

2. Chair: Robert Castiglione
Morrie Lipson and Peter Vallentyne, "Equal Opportunity
and the Family"
Commentator: Jeffrey Blustein

G. Special Panel Arranged by the APA Committee on the Teaching of Philosophy, *Richmond*

Topic: Teaching the Teaching of Philosophy

Chair: Merillee Salmon

Panel: Martin Benjamin, Baruch Brody, and Ernest Sosa

H. Special Panel Arranged by the APA Committee on the Status of Women, *Alexandria*

Chair: Alison M. Jaggar, Report on Work of the Committee Topic: Feminism and Racism

I. Maria Lugones, "The Logic of Pluralism"

II. Adrian Piper, "Higher Order Discrimination"

III. Elizabeth V. Spelman, "Privilege and Plethoraphobia"

IV. Paula Rothenberg, "Constructing, Deconstructing, and Reconstructing Difference"

APA PROGRAM

Digitized by Arya Samaj Foundation Chennai and eGangotri Group Meetings (See Appendix for Programs)

9:00 SARTRE CIRCLE, Calvert

9:00 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING

Woodley

9:00 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY

Marshall

11:15 RADICAL PHILOSOPHER'S ASSOCIATION AND SOCIETY FOR WOMEN IN PHILOSOPHY, *Arlington*

11:30 A.M.-1:30 P.M. BUSINESS MEETING, Sheraton Ballroom

AFTERNOON

2:15-4:15 Session III

A. Symposium: Plato and the Creation of the Socratic Dialogue

Baltimore

Chair: Rosemary Desjardins Speaker: Charles H. Kahn

Commentator: Charles L. Griswold, Jr.

B. Special Invited Panel: Philosophy in the Media

North Cotillion Ballroom

Chair: Virginia Held

Speakers: Fred Friendly, Douglas Kellner, and Alexander Nehamas

C. Book Discussion: Remembering: A Phenomenological Approach

Dover

Chair: Debra Bergoffen Author: Edward S. Casey

Commentators: Bernard Dauenhauer and David Carr

D. Book Discussion: Harmless Wrongdoing, Wilmington

Chair: Ferdinand Schoeman

Author: Joel Feinberg

Commentators: Douglas Husak and Judith Jarvis Thomson

E. Colloquium: Descartes and his Legacy, Richmond

1. Chair: Marjorie Grene

C. A. Bonnen and Daniel Flage, "Descartes's Factitious Ideas"

Commentator: Jeffrey Tlumak

2. Chair: Margaret Wilson

Guenter Zoeller, "From Innate to A Priori: Aspects of a Comparison Between Leibniz and Kant"

Commentator: Richard Kennington

Digitized by Arya Samaj Foundation Chennai and eGangotri Thursday, December 29, Afternoon, Session III

F. Colloquium: Thomas Hobbes, Rockville

1. Chair: Jacques Taminiaux Rosamond Rosenberg, "Duties and Rights: A Hobbesian Response"

Commentator: Thomas Lott

2. Chair: Robert Ginsberg

Neil Berlad, "Hobbes's Glorious Citizen" Commentator: Joseph Filonowicz

G. Colloquium: American Philosophy, Alexandria

1. Chair: Stanley Thayer

Mark Mendell, "Dewey and the Logic of Legal Reasoning" (The William James Prize Paper)

Commentator: Larry Hickman

2. Chair: Eugene Fontinell

Charlene H. Seigfried, "James' Natural History Methodology"

Commentator: Ignas Skrupskelis

H. Colloquium: Truth and Reference, Annapolis

1. Chair: Douglas Joel Butler Nick Zangwill, "Conceptions of Truth and Indirect Contexts"

Commentator: Thomas Upton

2. Chair: Mark E. Richard

Paul Berckmans, "On Demonstrative Utterances" Commentator: Ann Bezuidenhout

I. Special Panel Arranged by the APA Committee on Teaching Philosophy at Two-year Colleges, Idaho

Topic: The Introduction to Philosophy Course

Moderator: Cathleen Johnson Wu

Panel: Marlene Carpenter, Clyde Ebenreck, Donald Gregory, and Steven Tanner

J. Special Colloquium Arranged by the APA Committee on Blacks in Philosophy, Holmes

Topic: The Philosophy of Alain Locke

Chair: Johnny Washington

Leonard Harris, "The Critical Philosophy of Alain Locke" Commentators: Thomas Slaughter and Jeffrey Stewart

Digitized by Arya Samaj Foundation Chennai and eGangotri Group Meetings (See Appendix for Programs)

- 2:15 SOCIETY FOR BUSINESS ETHICS, Arlington
- 2:15 NORTH AMERICAN NIETZSCHE SOCIETY, South Cotillion Ballroom
- 2:15 SØREN KIERKEGAARD SOCIETY, Colorado
- 2:15 SOCIETY FOR PHILOSOPHY AND PUBLIC AFFAIRS, Vermont
- 2:15 ASSOCIATION FOR PHILOSOPHY AND LIBERATION, Calvert
- 4:30 PRESIDENTIAL ADDRESS, Sheraton Ballroom
 Introduction by Nicholas Rescher
 Richard J. Bernstein, "Pragmatism, Pluralism, and the
 Healing of Wounds"

EVENING (6:30-9:30)

Group Meetings (See Appendix for Programs)

- 6:30 NORTH AMERICAN KANT SOCIETY, Richmond
- 6:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, Arlington
- 6:30 society for the philosophical study of genocide and the holocaust, Alexandria
- 6:30 SOCIETY FOR SYSTEMATIC PHILOSOPHY, Baltimore
- 6:30 SOCIETY FOR WOMEN IN PHILOSOPHY, Annapolis
- 6:30 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY, *Rockville*
- 6:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, Colorado
- 6:30 society for the philosophy of creativity, Idaho
- 6:30 SOCIETY FOR PHILOSOPHERS AT WORK IN THE WORLD, Vermont
- 6:30 WORLD PHENOMENOLOGY INSTITUTE, Marshall
- 6:30 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT, Calvert
- 6:30 ASSOCIATION FOR THE PHILOSOPHY OF EDUCATION, Woodley
- 6:30 JASPERS SOCIETY OF NORTH AMERICA, Wilmington
- 6:30 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE, Holmes
- 6:30 SOCIETY FOR THE STUDY OF MEDIEVAL AND RENAISSANCE PHILOSOPHY, Warren
- 9:00 RECEPTION, Sheraton Ballroom

FRIDAY, DECEMBER 30

9:00 A.M.-12:00 Noon REGISTRATION, Registration Desk, Lobby Level 9:00 A.M.-2:00 P.M. BOOK DISPLAYS, Exhibit Hall A

PLACEMENT INFORMATION

For interviewers, 8:00 A.M.-12:00 Noon, Atrium #2 For candidates, 9:00 A.M.-12:00 Noon, Exhibit Hall C

9:00 A.M. EXECUTIVE COMMITTEE MEETING, Congressional

Friday, December 30

MORNING

9:00-11:00 Session IV

A. Symposium: Argument and Perception, Wilmington

Chair: Alan Gewirth

Speaker: Michael R. DePaul Commentator: R. B. Brandt

B. Symposium: What is Still Valuable in Husserl's Analysis of Inner

Time-consciousness, North Cotillion Ballroom

Chair: Charles Sherover Speaker: Peter K. McInerney Commentator: J. Claude Evans

C. Symposium: Feminist Skepticism, South Cotillion Ballroom

Chair: Sandra Harding Speaker: Susan Bordo

Commentator: Naomi Scheman

D. Book Discussion: Psychosemantics, Dover

Chair: Jaegwon Kim Speaker: Jerry Fodor

Commentators: Lynn Rudder Baker and Paul A. Boghossian

E. Colloquium: Epistemology and Social Criticism, Richmond

1. Chair: Andrew Buchwalter Nan Nan Lee, "Adorno's Conception of Critique" Commentator: Kenneth Baynes

2. Chair: Jorge Garcia Lenore Langsdorf, "Epistemic Man and Emergent Self" Commentator: Richard Schacht

F. Colloquium: Berkeley, Baltimore

1. Chair: Richard Talaska Phillip D. Cummins, "Berkeley's Unstable Metaphysics" Commentator: Harry Bracken

2. Chair: Karen Hanson Donald Baxter, "Berkeley, Perception, and Identity" Commentator: Robert Oakes

G. Colloquium: Responsibility, Annapolis

1. Chair: William Gerber Jean Rumsey, "Responsibility for Character" Commentator: Susan Wolf

2. Chair: Marcia Lind Larry May, "Shared Responsibility" Commentator: Patricia Greenspan

H. Colloquium: Aristotle, Rockville

1. Chair: Allan Gotthelf

Helen S. Lang, "Aristotle and Philoponus on Things Which Are"

Commentator: Andrea Croce Birch

2. Chair: Ronna Burger
James Martin, "Active Mind and the Text of De Anima 3:5"
Commentator: Daniel Devereux

11:15 A.M.-12:15 P.M. BUSINESS MEETING (Second Session, if needed), Sheraton Ballroom

AFTERNOON

1:30-4:30 Session V

A. Symposium: Law and Society, Sheraton Ballroom Chair: Richard Bernstein

Speaker: Jacques Derrida

Commentator: Thomas A. McCarthy

B. Symposium: Individuation and Self-knowledge, Wilmington

Chair: Joseph Camp Speaker: Tyler Burge

Commentator: Donald Davidson

C. Invited Papers:

1. The Philosophical Implications of Quantum Mechanics

Dover

Chair: Paul Durbin Speaker: David Albert

Commentator: Nancy Cartwright

2. Africana Philosophy
Chair: Frank Kirkland
Speaker: Lucius Outlaw
Commentator: John Murungi

D. Colloquium: Epistemology II, Richmond

1. Chair: Martha Bolton
Lois Frankel, "Hows and Whys: Causation Unlocked"
Commentator: Jonathan Kvanvig

2. Chair: Jonathan Adler Casey Swank, "Why I Am Not a Reliabilist" Commentator: Alvin Goldman

3. Chair: Jane Duran
Arthur Falk, "Is C. I. Lewis' Epistemology Bayesian?"
Commentator: Davis Baird

Friday, December 30, Afternoon, Session V

E. Colloquium: Philosophy of Mind, Baltimore

1. Chair: Michael Tye
Anne Jaap Jacobson, "How Folk Psychology is Not Causal
Theory"

Commentator: Dorit Bar-On

2. Chair: Georges Rey James S. Kelly, "Data before Theory" Commentator: Susan J. Brison

3. Chair: George Wilson
Bredo C. Johnsen, "Mental States as Mental"
Commentator: Francis Egan

F. Colloquium: Rights and Legal Foundations, Annapolis

Chair: Diana T. Meyers
 Ruth Lucier, "Inalienable Rights: A Plea for Open Options"
 Commentator: Judy Lichtenberg

2. Chair: Baine Harris
Frederick Kellogg, "Pursuing Legal Foundations"
Commentator: Christopher Gray

3. Chair: Richard Barber
Marilyn Friedman, "Friendship, Choice and Change"
Commentator: Laurence Thomas

Group Meetings (See Appendix for Programs) 1:30 AMERICAN SOCIETY FOR VALUE INQUIRY, *Arlington* 1:30 RADICAL PHILOSOPHER'S ASSOCIATION, *Colorado*

APA PROGRAM PARTICIPANTS

Digitized by Arya Samai Foundation Chennai and eGangotri PROGRAM PARTICIPANTS EIGHTY-FIFTH ANNUAL MEETING

ACKERMANN, Robert, University of Massachusetts/Amherst	I-B
ADAMS, Marilyn McCord, University of California/Los Angeles	II-B
ADLER, Jonathan, City University of New York/Brooklyn College	V-D
ALBERT, David, Columbia University	V-D V-C
ALCOFF, Linda, Kalamazoo College, Michigan	I-D
AMBROSIO, Francis, Georgetown University	I-D
BABICH, Babette, Denison University	I-D
BAIRD, Davis, University of South Carolina/Columbia	V-D
BAKER, Lynn Rudder, Middlebury College, Vermont	IV-D
BALSLEV, Anindita	I-I
BARBER, Richard, University of Louisville	V-F
BAR-ON, Dorit, University of Rochester	V-E
BAXTER, Donald, Princeton University	IV-F
BAYNES, Kenneth, State University of New York/Stony Brook	IV-E
BENJAMIN, Martin, Michigan State University	II-G
BERCKMANS, Paul, University of North Carolina/Chapel Hill	III-H
BERGOFFEN, Debra, George Mason University	III-C
BERLAD, Neil, Outremont, Canada	III-F
BERNSTEIN, Richard J., Haverford College, Pennsylvania	V-A
Presidential	
BEZUIDENHOUT, Ann, University of Michigan/Ann Arbor	III-H
BIRCH, Andrea Croce, St. Mary's College, Minnesota	IV-H
BLUSTEIN, Jeffrey, Mercy College, New York	II-F
BOGHOSSIAN, Paul A., University of Michigan/Ann Arbor	IV-D
BOLTON, Martha, Rutgers University/New Brunswick	V-D
	III-E
BONNEN, C. A., University of Texas/Austin	IV-C
BORDO, Susan, Le Moyne College, New York BOYHILL Represent University of North Carolina (Chanel Hill	I-C
BRACKEN Harman McCill Harmanity Montreal Canada	IV-F
BRACKEN, Harry, McGill University, Montreal, Canada	IV-A
BRANDT, R. B., University of Michigan/Ann Arbor	V-E
BRISON, Susan J., New York University	II-G
BRODY, Baruch, Rice University	IV-E
BUCHWALTER, Andrew, University of North Florida	V-B
BURGE, Tyler, University of California/Los Angeles	IV-H
BURGER, Ronna, Tulane University	III-H
BUTLER, Douglas Joel, North Carolina State University	V-B
CAMP, Joseph, University of Pittsburgh	I-D
CAPUTO, Richard M., Assumption College, Massachusetts	II-A
CAPPENTED A CAPPEN	II-A
CARPENTER, Marlene, Prince George's Community College,	*** *
Maryland CARD Days and Cards	III-I
CARR, David, University of Ottawa, Canada	III-C
CASTY RICHT, Nancy, Stanford University	V-C
CASEY, Edward S., State University of New York/Stony Brook	III-C

Digitized by Arya Samaj Foundation Chennal and eGangotri CASTIGLIONE, Robert, Rhode Island College, Rhode Island	II_F
CHAKRAVARTY, Arindam	I-I
CHARTOPADHYAYA, D. P.	I-I
COBB-STEVENS, Richard, Boston College, Massachusetts	II-C
CODER, David, National Endowment for the Humanities	I-E
COHEN, Joshua, Massachusetts Institute of Technology	I-C
COOPER, John, Princeton University	II-D
CUMMINS, Phillip D., University of Iowa	IV-F
CURZER, Howard, Texas Tech University	II-D
DALLERY, Arlene B., La Salle University	I-H
DAUENHAUER, Bernard, University of Georgia	III-C
DAVIDSON, Donald, University of California/Berkeley	V-B
DE CEW, Judith, Clark University	I-F
DePAUL, Michael R., University of Notre Dame	IV-A
DERRIDA, Jacques, École des Hautes Études et Sciences Sociales,	11-11
Paris	V-A
DESJARDINS, Rosemary, Haverford College, Pennsylvania	III-A
DEUTSCH, Eliot, University of Hawaii/Manoa	I-I
DEVEREUX, Daniel, University of Virginia	IV-H
DONAGAN, Alan, California Institute of Technology	I-A
DOPPELT, Gerald, University of California/La Jolla	I-C
DOUBLE, Richard, Purdue University	I-F
DROST, Mark P., University of Rochester	II-C
DURAN, Jane, University of California/Santa Barbara	V-D
DURBIN, Paul, University of Delaware	V-C
EBENRECK, Clyde, Prince George's Community College, Maryland	III-I
EGAN, Francis, Rutgers University/New Brunswick	V-E
EVANS, J. Claude, Washington University	IV-B
FALK, Arthur, Western Michigan University	V-D
FEINBERG, Joel, University of Arizona	III-D
FELDMAN, Richard, University of Rochester	I-E
FILONOWICZ, Joseph, Long Island University	III-F
FLAGE, Daniel, University of Texas/Austin	III-E
FLYNN, Thomas, Emory University	I-A
FODOR, Jerry, City University of New York (Craduate C	IV-D
Torvince L., Eugene, City University of New York Queens College	III-G
2) Zois, Chiversity of Colorado	V-D
FRASER, Nancy, Northwestern University	I-H
FREDE, Michael, Princeton University	I-A
FRIEDMAN, Marilyn, Bowling Green State University	V-F
TRIEDLI, Fled, Columbia University	III-B
FUMERION, Richard, University of Iowa	I-E
GALISON, Peter, Stanford University	I-B
GANGADEAN, Ashok, Haverford College Pennsylvesis	I-I
TOTAL TOTAL TELLIDETOWN I INTROMOTER	IV-E
GASCHE, Rodolphe, University of Posher	II-A
GAUTHIER, David, University of Pittsburgh	I-C

APA PROGRAM PARTICIPANTS	585
Digitized by Arya Samaj Foundation Chennai and eGangotri	
GERBER, William, Washington, DC	IV-C
GEWIRTH, Alan, University of Chicago	IV-A
GIBSON, Mary, Rutgers University/New Brunswick	I-H
GINSBERG, Robert, Pennsylvania State University	III-F
GLEB, Gary, Rutgers University/New Brunswick	I-E
GOLDMAN, Alvin, University of Arizona	V-D
GOTTHELF, Allan, Trenton State College, New Jersey	IV-H
GRAY, Christopher, Concordia University, Canada	
GREENSPAN, Patricia, University of Maryland	V-F
GREGORY, Donald, Northern Virginia Community College	IV-G
GRENE, Marjorie, Virginia Polytechnic Institute and State University	III-I
GRISWOLD, JR., Charles L., Howard University	
HACKING, Ian, University of Toronto	III-A
HAMPTON, Jean, University of Pittsburgh	I–B
HANSON, Karen, Indiana University/Bloomington	I-C
HAPDIMON Michael Massachusette Institute C.T. 1	IV-F
HARDING Sandra University of Dela	II-F
HARDING, Sandra, University of Delaware	IV-C
HARMAN, Gilbert, Princeton University	I–F
HARRIS, Baine, Old Dominion University	V-F
HARRIS, Leonard, Morgan State University	III-J
HATAB, Lawrence, Old Dominion University	I-D
HEELAN, Patrick A., State University of New York/Stony Brook	I–B
HEFFERNAN, George, University of Notre Dame	II-C
HELD, Virginia, City University of New York/Hunter College	III-B
HENSON, Richard, Rutgers University/New Brunswick	I-F
HETHERINGTON, Stephen C., West Virginia University	I-E
HICKMAN, Larry, Texas A & M University	III-G
HINTIKKA, Jaakko, Florida State University	I-G
HOLLAND, Monica, Indiana University/Bloomington	I-E
HUSAK, Douglas, Rutgers University/New Brunswick	III-D
JACOBSON, Anne Jaap, Rutgers University/New Brunswick	V-E
JAGGAR, Alison M., University of Cincinnati	II-H
JOHNSEN, Bredo C., University of Houston/University Park	V-E
KAHN, Charles H., University of Pennsylvania	III-A
KELLNER, Douglas, University of Texas/Austin	III-B
KELLOCC Frederick Westington DC	
KELLOGG, Frederick, Washington, DC	V-F
KELLY, James S., Elmhurst College, Illinois	V-E
KELLY, Michael, Journal of Philosophy/Columbia University	I-D
KENNINGTON, Richard, Catholic University of America	III–E
KIM, Jaegwon, University of Michigan/Ann Arbor	IV-D
KIRKLAND, Frank, City University of New York/Hunter College	V-C
KORNBLITH, Hilary, University of Vermont	I-E
AVAINVIG, Ionathan, Texas A & M University	V-D
Helen S., Trinity College, Connecticut	IV-H
ANGSDORF, Lenore, University of Texas/Arlington	IV-E
Nan Nan Chicago Illinois	IV-E
LEFTOW, Brian, Fordham University	II-E

LELAND, Dorothy, Purdue University	I–H
LICHTENBERG, Judy, University of Maryland	V-F
LIND Marcia, University of Cincinnati	IV-G
LIPSON, Morrie, University of Western Ontario, Canada	II-F
LOTT, Thomas, University of Massachusetts/Boston	III-F
LUCIER, Ruth, Bennett College, North Carolina	V-F
LUGONES, Maria, Carleton College, Minnesota	II-H
MADIGAN, Arthur, Boston College, Massachusetts	II-E
MANN, Jesse, Georgetown University	I-F
MARSHALL, John, University of Virginia	I-F
MARTIN, James, University of Wyoming	IV-H
MATILAL, Bimal Krishna	I-I
MAY, Larry, Purdue University	IV-G
McCARTHY, Thomas A., Northwestern University	V-A
McINERNEY, Peter K., Oberlin College, Ohio	IV-B
McPHERRAN, Mark L., University of Maine/Farmington	II-D
MENDELL, Mark, University of Pennsylvania	III-G
MEYERS, Diana T., University of Connecticut/Storrs	V-F
MILLER, Izchak, University of Pennsylvania	II-C
MOHANTY, Jitendranath, Temple University	I-I
MURUNGI, John, Towson State University	V-C
NEHAMAS, Alexander, University of Pennsylvania	III-B
OAKES, Robert, University of Missouri	IV-F
OUTLAW, Lucius, Haverford College, Pennsylvania	V-C
PAPPU, S. S. Rama Rao, Miami University	I-I
PETERMAN, John E., William Paterson College, New Jersey	II-E
PIPER, Adrian, University of California/San Diego	II-H
POTTER, Karl H., University of Washington	I-I
POTTER, Vincent, Fordham University	I-G
PUTNAM, Hilary, Harvard University	I-G
RASMUSSEN, Douglas, St. John's University New York	II-D
RAVVEN, Heidi M., Hamilton College, New York	II-F
RESCHER, Nicholas, University of Pittsburgh	
Introduction: Providential	Address
KE1, Georges, University of Maryland	V-E
RICHARD, Mark E., Tufts University	III-H
ROBIN, Richard S., Mount Holyoke College Man	I-G
202, Inicia, I lince George's Community Call	
ROSENBERG, Rosamond, City University of New York/	I-H
ROSS James University of D	III-F
ROSS, James, University of Pennsylvania	II-B
ROTH, Michael D., Franklin and Marshall College, Pennsylvania	I-E
	II-H
, Johns Charlott Childership Of Donnard	IV-G
of Dittal	II-G
SCHACHT, Richard, University of Illinois/Urbana-Champaign SCHEMAN, Naomi University of Mineral Champaign	IV-E
SCHEMAN, Naomi, University of Minnesota	IV-C

Digitized by Arya Samaj Foundation Chennai and eGangotri	587
SCHOEMAN, Ferdinand, University of South Carolina/Columbia	III-D
SCHOLLMEIER, Paul, George Washington University	II-D
SCOTT, Charles, Vanderbilt University	II-A
SEIGFRIED, Charlene H., Purdue University	III-G
SEIGFRIED, Hans, Loyola University of Chicago	
SHEROVER, Charles, City University of New York/Hunter College	I-D
SHOPE, Robert, University of Massachusetts/Boston	IV-B
SILVERMAN, Allan, Ohio State University	I-E
SIMPSON, Lorenzo, James Madison University	II-E
SKRUPSKELIS, Ignas, University of South Carolina/Columbia	I-D
SLAUGHTER, Thomas, Bentley University	III-G
SLOTE, Michael, University of Maryland	III-J
SMITH, Nicholas, Virginia Polytechnic Institute & State University	I-F
SMITH, Robin, Kansas State University	II-E
SOKOLOWSKI, Robert, Catholic University of America	II-D II-C
SOSA, Ernest, Brown University	II-G
SPELMAN, Elizabeth V., Smith College, Massachusetts	II-H
STEPELEVICH, Lawrence, Villanova University	II-F
STEWART, Jeffrey, Anacostia Museum	
STRANGE, Stephen, University of Pittsburgh	III-J II-E
SWANK, Casey, University of Alabama	
TALASKA, Richard, Xavier University	V-D IV-F
TAMINIAUX, Jacques, Boston College, Massachusetts	III-F
TANNER, Steven, Dundalk Community College	III-I
THAYER, Stanley, City University of New York/Graduate Center	III-G
THOMAS, Laurence, Oberlin College, Ohio	V-F
THOMSON, Judith Jarvis, Massachusetts Institute of Technology	III-D
TLUMAK, Jeffrey, Vanderbilt University	III-E
TRAGESSER, Robert, Barnard College/Columbia University	II-C
TYE, Michael, Temple University	V-E
UPTON, Thomas, Gannon University	III-H
VALLENTYNE, Peter, University of Western Ontario, Canada	II-F
WALKER, Margaret, Fordham University	I–F
WARNKE, Georgia, Yale University	I-D
WASHINGTON, Johnny, Florida Atlantic University	III-J
WILES, Ann M., James Madison University	I-F
WILSON, George, Johns Hopkins University	V-E
WILSON, Margaret, Princeton University	III-E
WIPPEL, John, Catholic University of America	II-B
WIREDU, Kwasi, University of South Florida	I-C
WOLF, Susan, Johns Hopkins University	IV-G
WU, Kathleen Johnson, University of Alabama	III-I
YOUNG, Iris M., Worcester Polytechnic Institute, Massachusetts	I-H
ZANGWILL, Nick, East Carolina University, North Carolina	III-H
ZOELLER Guenter University of Iowa	III_F

THE JOURNAL OF PHILOSOPHY

Digitized by Arya Samaj Foundation Chennai and eGangotri

APPENDIX: PROGRAMS OF GROUP MEETINGS

TUESDAY, DECEMBER 27

I. EVENING (8:00-11:00)

1. 8:00 NORTH AMERICAN DIVISION OF THE SCHOPENHAUER SOCIETY

Richmond

Topic: Jubilaeumskongress Chair: Eric von der Luft

- I. David Cartwright, "Schopenhauer: 1788-1988"
- II. Ivan Soll, "Desires and Their Gratification"
- III. Joachim Baer, "Schopenhauer and Russian Literature"
- IV. John Atwell, "The World as Representation: Its Two Meanings"
- 2. 8:00 CONFERENCE OF PHILOSOPHICAL SOCIETIES, Arlington
 Topic: The Annual Business Meeting and General Assembly of
 Philosophical Societies
- 3. 8:00 SOCIETY OF CHRISTIAN PHILOSOPHERS, *Dover*Topic: Romans 1:20 and the Possibility of a Natural Theology

 Chair: Robert Merrihew Adams
 - I. Norris Clarke, "Thomas Aquinas on Romans 1:20 and the Possibility of a Natural Theology"
 - II. Arvin Vos, "John Calvin on *Romans* 1:20 and the Possibility of a Natural Theology"

Commentator: Kenneth Konyndyk

4. 8:00 SOCIETY FOR CLASSICAL REALISM, Baltimore

Topic: Realism and Commitment to Values in a True

Philosophical Pluralism

Chair and Commentator: Mary Carman Rose

- I. Fran Gillespie, "Commitment to Values in Education: The Contemporary and Future Relevance of the Ignatian Ideal"
- II. Fritz Wenisch, "A Realist Foundation of Ethics"
- 5. 8:00 LENSGRINDERS' ASSOCIATION, Marshall Reception for Job Seekers

WEDNESDAY, DECEMBER 28

II. MORNING (9:00-11:00)

1. 9:00 INTERNATIONAL ST. THOMAS SOCIETY, Arlington
Chair: Mary Rose Barral
James F. Keenan, "The Unity of the Virtues: Charity or
Prudence, Goodness or Rightness?"

APA PROGRAM, APPENDIX
Digitized by Arya Samaj Foundation Chennai and eGangotri

2. 9:00 SOCIETY FOR WOMEN IN PHILOSOPHY, Alexandria

I. In Honor of Distinguished Women Philosophers

Chair: Mary Mothersill Honoree: Leigh Cauman

Leigh Cauman, "On Supposition: Pedagogical Problems in Logic Teaching"

Commentators: Sue Larson and Ti-Grace Atkinson

II. Business Meeting Chair: Jana Sawicki

3. 9:00 SOCIETY FOR PHILOSOPHY AND TECHNOLOGY AND SOCIETY FOR

THE STUDY OF PROCESS PHILOSOPHY, Dover

Topic: Philosophy of Technology by Frederick Ferré

Chair: Paul T. Durbin

Speakers: Joseph Pitt and Peter Limper

Commentator: Frederick Ferré

4. 9:00 NORTH AMERICAN NIETZSCHE SOCIETY, Wilmington Chair: Gary Shapiro

I. Symposium: Perspectives on Perspectivism

a. Robin Alice Roth, "Nietzsche's Metaperspectivism"

b. Babette B. Babich, "Nietzsche and the Philosophy of Scientific Power"

c. Daniel Conway, "Beyond Realism: Nietzsche's 'New Infinite'"

II. Richard Shusterman, "Nietzsche, Nehamas, and Organic Unity"

Commentator: Alexander Nehamas

III. Business Meeting (Chair: Richard Schacht)

5. 9:00 WORLD PHENOMENOLOGY INSTITUTE, Baltimore

Topic: Recent American Phenomenology: Anna-Teresa Tymieniecka's Logos and Life

Chair: Marlies Kronegger

I. Dallas Laskey, "A New Departure for the Critique of Reason: Phenomenology of the Creative Act"

II. Leo Rauch, "Logos and the Human Condition in the Thought of Anna-Teresa Tymieniecka"

6.9:00 international philosophers for the prevention of

NUCLEAR OMNICIDE, Vermont
Topic: Perestroika: Harbinger of World Peace?

Chair: George Hampsch

I. Howard Parsons, "Perestroika and the Psychological Revolution: The Human Factor"

II. Paul Allen, III, "Perestroika, Peace, and the New Openness"

III. Yuri Zamoshkin

Wednesday, December 28, Morning, II

7. 9:00 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING Woodley

Chair: Donald Hatcher

I. Dale Jacquette, "The Hidden Logic of Slippery Slope Arguments"

Commentator: Roy Sorensen

II. Jaakko Hintikka, "What Is the Relation of Informal Logic to Deductive Logic?"

Commentator: Gilbert Harman

8. 9:00 SOCIETY FOR CLASSICAL REALISM, Annapolis

Topic: History, Analysis, and Creativity in the Burgeoning of Realism

Chair and Commentator: Mary Carman Rose

- Richard Geraghty, "Why Realism Demands Immateriality in the Knower"
- II. Raymond J. Langley, "Mirror, Mirror . . . Who is the Real Self?"

III. LATE MORNING (11:15-1:15)

1. 11:15 AMERICAN SOCIETY FOR VALUE INQUIRY, Richmond

Topic: Presidential Address

Chair: Sander Lee

- I. Lisa H. Newton, "Inquiries into Practice: Problems and Perspectives"
- II. Business Meeting
- 2. 11:15 SOCIETY FOR THE PHILOSOPHICAL STUDY OF GENOCIDE AND THE HOLOCAUST AND THE APA COMMITTEE ON BLACKS IN PHILOSOPHY, *Arlington*

Topic: Group Identity in Genocide and Holocaust

Chair: Alan Rosenberg

- I. Laurence Thomas, "Blacks, Jews, and Group Autonomy"
- II. Robert S. Gottlieb, "Political and Personal Uses of Histories of Suffering"
- 3. 11:15 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, Colorado
- 4. 11:15 INTERNATIONAL SOCIETY FOR CHINESE PHILOSOPHY, Warren Chair: Lik Kuen Tong
 - I. Ellen May Chen, "Taoist Legacy to the New World Religious Consciousness"
 - II. Kwang-Sae Lee, "Confucianism and Kant"
 - III. Sung Bae Park, "Chinul and Toegye: A Variation on Buddhism and Neo-Confucianism in Korea"

APA PROGRAM, APPENDIX

Digitized by Arya Samaj Foundation Chennai and eGangotri

5. 11:15 SARTRE CIRCLE, Calvert

Chair: Wilfrid Desan

I. Thomas R. Flynn, "Sartre and the Poetics of History"

Commentator: Peter Caws

II. William McBride, "Sartrean Ethics and Political Theory in the Light of Volume II of the Critique of Dialetical Reason"

Commentator: Sonia Kruks

6. 11:15 JASPERS SOCIETY OF NORTH AMERICA, Annapolis

Chair: Robert S. Corrington

I. Gita S. van Heerden, "Epochal Consciousness: A
Postmodern Response to the Spiritual Situation of the
Time"

Commentator: Harold A. Durfee

II. Robert L. Bock, "Jaspers and Politics"

Commentator: Raymond J. Langley

IV. AFTERNOON (2:00-5:00)

1. 2:00 SOCIETY FOR SYSTEMATIC PHILOSOPHY, Arlington

Topic: Morality and Community

I. Robert C. Neville, "Morality with Community"

II. Richard Dien Winfield, "Morality without Community"

III. Terry Pinkard, "Hegelian Community"

2. 2:00 CAUCUS OF DEPARTMENT CHAIRS, Vermont

Topic: Assessment and Accountability: A Panel Discussion

Chair: Jay F. Rosenberg

Panel: Theodore M. Benditt, Myles Brand, and Michael Hooker

3. 2:00 AMERICAN ASSOCIATION FOR THE PHILOSOPHIC STUDY OF SOCIETY, Woodley

Topic: Loren E. Lomasky's Persons, Rights, and the Moral Community

Chair: Douglas J. Den Uyl

Speakers: Tibor R. Machan, Eric Mack, Christopher W. Morris, and Douglas B. Rasmussen

Commentator: Loren E. Lomasky

V. LATE AFTERNOON (5:15-7:15)

1. 5:15 society for the philosophical study of marxism, Richmond

Topic: Marxism and Law

Chair: John Ryder

Terry DiFilippo, "Marxism and Legality"

Wednesday, December 28, Late Afternoon, V

2. 5:15 SOCIETY FOR WOMEN IN PHILOSOPHY, Alexandria

Topic: Feminist Reflections on Social Change

Chair: Jana Sawicki

- I. Linda Singer, "Making a Mess: Symbolism, Power, and Social Change"
- II. Shere Hite
- III. Dion Farquhar, "Fiction and Strategies of Social Change"
- 3. 5:15 WORLD PHENOMENOLOGY INSTITUTE, Baltimore

Chair: Evelyn Barker

- I. Ynhui Park, "Individualization and the Unity of Everything-there-is-alive: Traditional Ontology Transformed"
- II. Mary Rose Barral, "Introducing the Phenomenology of Life and of the Human Being of Anna–Teresa Tymieniecka
- 4. 5:15 SOCIETY FOR ANCIENT GREEK PHILOSOPHY, Colorado Chair: Anthony Preus
 - I. Scott Austin, "The Ignorance of Socrates"
 - II. Donald Morrison, "Ontological Structure in Aristotle"
- 5. 5:15 AYN RAND SOCIETY, Idaho

Topic: Perceptual Appearance: Realism versus

Representationalism

Chair: George V. Walsh Speaker: David Kelly

Commentator: Jaegwon Kim

6. 5:15 CHARLES S. PEIRCE SOCIETY, Kansas

Chair: Andrew J. Reck

- I. Richard A. Smyth, "The Beginning of 'Some Consequences'"
- II. Peirce Society Essay Contest Award Paper
- 7. 5:15 SOCIETY FOR ASIAN AND COMPARATIVE PHILOSOPHY, Vermont Topic: Emotion in Comparative Perspective
 - I. Chair: Joel Marks (Indian Tradition)
 Anindita Niyogi Balskev, "Asmitā as a Kleśa, or the I-sense as an Emotion"

Commentator: Jitendranath Mohanty

II. Chair: Antonio S. Cua (Chinese Tradition)

David Wong, "Is There a Distinction between Reason and Emotion in Confucianism?"

Commentator: Craig K. Ihara

III. Chair: Glen Martin (Japanese Tradition)
Graham Parkes, "The Transmutation of Emotion in Nietzsche and Zen"

Commentator: Joan Stambaugh

APA PROGRAM, APPENDIX

Digitized by Arya Samaj Foundation Chennai and eGangotri

8. 5:15 INTERNATIONAL BERKELEY SOCIETY, Calvert

Topic: Berkeley

Chair: Kenneth P. Winkler

9. 5:15 PERSONALISTIC DISCUSSION GROUP, Rockville

Topic: Personalism: New Directions

Chair: Thomas Buford

Speakers: John Lavely and Erazim Kohak

10. 5:15 BERTRAND RUSSELL SOCIETY, Holmes

Chair: David E. Johnson

Jan Dejnozka, "The Ontological Foundations of Russell's

Theory of Modality"

Commentator: Stewart Umphrey

11. 5:15 G. E. MOORE SOCIETY, Annapolis

Topic: Moore and Wittgenstein

Chair: Dennis Rohatyn

Speakers: Garth Hallett, Nicholas Rescher, and Kathryn Doran

12. 5:15 ASSOCIATION FOR THE PHILOSOPHY OF THE UNCONSCIOUS

Marshall

Topic: A Critical Look at Lacan

Chair: Wilfried ver Eecke

I. Devra Simiu, "A Critical Look at Lacan's Mirror Stage Theory"

Commentator: Richard Boothby

II. Debra Bergoffen, "Lacan's Hamlet: Mourning, Woman, and the Phallus"

13. 5:15 AMERICAN ASSOCIATION OF PHILOSOPHY TEACHERS, Arlington

Chair: Rosalind Ekman Ladd

Gareth Matthews, "Teaching Philosophy as Reconstructing

Childhood"

Commentator: Karen J. Warren

VI. EVENING (7:30-10:30)

1. 7:30 ASSOCIATION FOR THE PHILOSOPHY OF EDUCATION, Richmond

Topic: Richard Rorty and Education

Chair: Sophie Haroutunian-Gordon

I. Rene V. Arcilla, "Edification, Conversation, and

Narrative: Rortyan Motifs for Philosophy of Education"
II. Carol Nicholson, "Lyotard and Rorty on Postmodern

Education"

Commentator: Richard Rorty

Wednesday, December 28, Evening, VI

2. 7:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, Arlington

Topic: Organizing New Societies

Chair: John Abbarno

Speakers: Sander Lee (The American Society for Value Inquiry), Nance Cunningham Butler (The Society for Philosophers at Work in the World), and William McBride (Sartre Circle)

3. 7:30 NORTH AMERICAN DIVISION OF THE SCHOPENHAUER SOCIETY

Alexandria

Chair: David Cartwright

I. Steven Neeley, "A Critical Note on Schopenhauer's Concept of Human Salvation"

II. Honi Haber, "Taking Seriously the View That All is One"

III. Richard McCarty, "Schopenhauer on Moral Worth"

IV. James Snow, "Schopenhauer, Kant, and the Possibility of Metaphysics"

4. 7:30 SANTAYANA SOCIETY, Baltimore

Chair: Paul G. Kuntz

Topic: Presentation of the Critical Edition of The Sense of Beauty Presenters: Betty Stanton, Herman J. Saatkamp, Jr., and William G. Holtzberger

Speaker: Arthur C. Danto Commentator: Willard Arnett

5. 7:30 RADICAL PHILOSOPHER'S ASSOCIATION, Colorado

I. Bill Martin, "The Office(s) of Philosophy: An Institutional Critique of the Theoretical Plane"

Commentator: Thomas Simon

6. 7:30 LEIBNIZ SOCIETY OF AMERICA, Idaho

Chair: Margaret Wilson

I. Robert Merrihew Adams, "Form and Matter in Leibniz"

II. Business Meeting

7. 7:30 GANDHI-KING SOCIETY, Calvert

Chair: Robert Reuman

Topic: Nonviolence and the Just War

Panel: Ronald E. Santoni (Leader and Discussant), Douglas Lackey, Kenneth Kemp, and Kristin Schrader– Frechette

8. 7:30 SOCIETY FOR THE STUDY OF MEDIEVAL AND RENAISSANCE PHILOSOPHY, *Annapolis*

Chair: Aryeh L. Motzkin

I. Calvin Normore, "Accidental Necessity: A Short Medieval History"

II. Edward P. Mahoney, "Ethical Aspects of the Great Chain of Being"

APA PROGRAM, APPENDIX

Digitized by Arya Samaj Foundation Chennai and eGangotri

- 9. 7:30 SOCIETY FOR THE STUDY OF WOMEN PHILOSOPHERS, Holmes
- 10. 7:30 INTERNATIONAL SOCIETY OF MEOPLATONIC STUDIES, Marshall Topic: Plotinus' Metaphysics

I. Eyjólfur Kjalar Emilsson, "Ennead IV.2: Some Comments and Questions"

II. Steven K. Strange, "Plotinus on Participation" Commentator: Michael Frede

THURSDAY, DECEMBER 29

VII. MORNING (9:00-11:00)

1. 9:00 SARTRE CIRCLE, Calvert

Chair: Phyllis S. Morris

I. Joseph S. Catalano, "Sartre's Concept of Authenticity"

II. Julien S. Murphy, "Sartre on the Side of Things"
Commentator: Linda Ann Bell

2. 9:00 ASSOCIATION FOR INFORMAL LOGIC AND CRITICAL THINKING

Woodley

I. Topic: Pragmatics and Informal Logic: The Work of Frans H. van Eemeren and Rob Grootendorst

Speakers: Anthony Blair and Lenore Langsdorf

Commentators: Rob Grootendorst and Frans H. van Eemeren

II. Business Meeting

3. 9:00 SOCIETY FOR THE ADVANCEMENT OF AMERICAN PHILOSOPHY

Marshall

Topic: Critical Perspectives on the Theory of Modernity

Chair: Kathleen Wallace Speaker: Lawrence Cahoone

Commentators: Sidney Gelber and Gary Shapiro

VIII. EARLY AFTERNOON (11:15-1:15)

1. 11:15 RADICAL PHILOSOPHER'S ASSOCIATION AND SOCIETY FOR WOMEN IN PHILOSOPHY, *Arlington*

I. Linda Alcoff, "The Unhappy Marriage of Feminism and Foucault"

Commentator: Kelly Oliver

II. Andrew McLaughlin, "Socialism and Ecology"

IX. AFTERNOON (2:15-4:15)

1. 2:15 SOCIETY FOR BUSINESS ETHICS, Arlington

Topic: An Alternative View of Business Ethics

Chair: W. Michael Hoffman

I. Edward F. McClennen, "Morality as a Public Good"

Commentator: J. Gregory Dees

II. Business Meeting

Thursday, December 29, Afternoon, IX

2. 2:15 NORTH AMERICAN NIETZSCHE SOCIETY, South Cotillion Ballroom Topic: Nietzsche and Allan Bloom's Nietzsche

Chair: Bernd Magnus

Panel: Werner Dannhauser, Alexander Nehamas, Richard Rorty, and Richard Schacht

3. 2:15 SOREN KIERKEGAARD SOCIETY, Colorado

Chair: John Donnelly

I. Steven M. Emmanuel, "Kierkegaard on Knowledge and Faith"

Commentator: David Wisdo

II. T. F. Morris, "Kierkegaard's Teleological Suspension of the Ethical"

Commentator: Earl McLane

III. Business Meeting

4. 2:15 SOCIETY FOR PHILOSOPHY AND PUBLIC AFFAIRS, Vermont

Chair: Diana T. Meyers

Anthony Weston, "Reconstruction in Applied Ethics: Beyond the Personhood Debate"

Commentators: Marilyn Friedman and Scott Lehman

5. 2:15 ASSOCIATION FOR PHILOSOPHY AND LIBERATION, Calvert Chair: Kate Lindeman

Michael D. Barber, "Phenomenology and the Looking Glass of Race"

Commentator: Joseph J. Godfrey

X. EVENING (6:30-9:30)

1. 6:30 NORTH AMERICAN KANT SOCIETY, Richmond

Chair: Arnulf Zweig

Topic: The Recently Discovered Kantian Fragment, "On Inner Sense": A Symposium

Panel: Hoke Robinson and Guenter Zoeller

2. 6:30 CONFERENCE OF PHILOSOPHICAL SOCIETIES, Arlington Topic: The World and American Philosophical Practice

Chair: John A. Loughney

Speakers: Fred G. Sturm, Robert C. Neville, John A. Loughney, and John E. Smith

3. 6:30 SOCIETY FOR THE PHILOSOPHICAL STUDY OF GENOCIDE AND THE HOLOCAUST, *Alexandria*

Topic: The Holocaust as Event

Chair: Sander Lee

I. Alan Rosenberg, "Was the Holocaust Unique? A Peculiar Question"

II. James Bernauer, "Nazi Ethics"

III. Business Meeting

Digitized by Arya Samaj Foundation Chennai and eGangotri

4. 6:30 SOCIETY FOR SYSTEMATIC PHILOSOPHY, Baltimore

Topic: Symposium on Conceiving Reality

I. Arindam Chakrabarti, "Idealist Refutations of Idealism"

Commentator: Edward Halper

II. Stephen Houlgate, "Thought and Reality in Kant and Hegel"

Commentator: Robert Berman

5. 6:30 SOCIETY FOR WOMEN IN PHILOSOPHY, Annapolis

Topic: Feminist Reviewing for Journals

Chair: Jeffner Allen (Hypatia)

Panel: Margaret A. Simons (Hypatia), Jean F. O'Barr (Signs), and Cynthia Enloe & Joni Seager (Women's Studies International Forum)

6. 6:30 american association for the philosophical study of society, Rockville

Topic: Aspects of Truth and Realism

- I. Barry Smith, "Realism and the Correspondence Theory of Truth"
- II. Andrew Melnyk, "Eliminativist Physicalism"
- III. Barbara Hannan, "Eliminativist Materialism Eliminated"
- 7. 6:30 SOCIETY FOR THE STUDY OF ETHICS AND ANIMALS, Colorado Chair: Lilly-Marlene Russow
 - I. Judith Barad, "Aquinas's Inconsistency on the Nature and Treatment of Animals"

Commentator: Daniel A. Dombrowski

II. Julie Dunlap, "The Adolescent as Environmental Ethicist" Commentator: Evelyn B. Pluhar

III. Laura Westra, "Ecology and Animals: Is There a Joint Ethic of Respect?"

Commentator: Harlan B. Miller

8. 6:30 SOCIETY FOR THE PHILOSOPHY OF CREATIVITY, Idaho
Topic: The New Thermodynamics: Creativity in Nature?
Chair: Paul G. Kuntz
Speaker: Pete A. Y. Gunter
Commentator: Ilya Prigogine

9. 6:30 SOCIETY FOR PHILOSOPHERS AT WORK IN THE WORLD, Vermont Topic: Timely Meditations: Issues facing Philosophers in Contemporary Society

Chair: Nance Cunningham Butler and Jean A. Campbell

Speakers: Brent Blackwelder, Ralph Dumain, and Patricia A.
Muoio

Digitized by Arya Samaj Foundation Chennai and eGangotri Thursday, December 29, Evening, \boldsymbol{X}

- 10. 6:30 WORLD PHENOMENOLOGY INSTITUTE, Marshall
 - I. Marlies Kronegger, "A New Phenomenology of Mind: Vindicating the 'Life Significance' of Literature"
 - II. Chung-ying Cheng, "The Confucian Approach to Creativity and the Phenomenological 'History of the Logos'"
 - III. Jadwiga Smith, "Creative Imagination and the Elemental Passions": A Challenge to the Current Theories of Literature"
- 11. 6:30 SOCIETY FOR IBERIAN AND LATIN AMERICAN THOUGHT, Calvert
 - I. John T. Graham, Ortega y Gasset, Pragmatism, and Historians
 - II. Ofelia Schutte, "Rereading Hegel's Master-Slave Dialectic in Latin America: Social Criticism by Zea, Freire, and Roig"
- 12. 6:30 ASSOCIATION FOR THE PHILOSOPHY OF EDUCATION, Woodley Topic: Education for the Critical Thinking of Intuition?

 Chair: Harvey Siegel
 - I. Donald Hatcher, "Is Critical Thinking Guilty of Reductionism?"
 - II. Hubert L. Dreyfus, "Teaching Intuition Rather than Critical Skills"
- 13. 6:30 JASPERS SOCIETY OF NORTH AMERICA, Wilmington

 Topic: World Congress Jaspers Sessions

 Panel: Alan Olson, George Pepper, Robert S. Corrington, Gregory Walters, William Kluback, Oswald Schrag, and Richard Howey

 Business Meeting
- 14. 6:30 SOCIETY FOR THE PHILOSOPHY OF SEX AND LOVE, Holmes
 Topic: Symposium on Roger Scruton's Sexual Desire
 Chair: Anthony F. Beavers
 I. Edward Johnson, "Scrutinizing Sexual Desire"
 II. Dan Sullivan, "Scruton on Sexual Desire"
 Commentator: David James
- 15. 6:30 SOCIETY FOR THE STUDY OF MEDIEVAL AND RENAISSANCE PHILOS-OPHY, Warren

Chair: James R. Long

- I. Seymour Feldman, "The Bible as an Epistemological Text: Some Remarks on Gersonides's Hermeneutics"
- II. Jorge Gracia, "John of St. Thomas on Individuation"

FRIDAY, DECEMBER 30°

XI. AFTERNOON (1:30-4:30)

1. 1:30 AMERICAN SOCIETY FOR VALUE INQUIRY, Arlington Topic: Philosophical Themes in Film

Chair: Thomas Magnell

I. Bat-Ami Bar On, "Platoon and the Protest of War"

II. Sander Lee, "The Role of Seeing in Lynch's Blue Velvet"

III. George J. Stack, "La Dolce Vita: Appearance and Reality, Innocence and Decadence"

2. 1:30 RADICAL PHILOSOPHER'S ASSOCIATION, Colorado

I. Joe Walsh, "Sartre, Marxism, and the Ethics of Revolutionary Violence"

Commentator: Anatole Anton

II. Business Meeting

THE JOURNAL OF PHILOSOPHY Digitized by Arya Samaj Foundation Chennai and eGangotri

GROUP PARTICIPANTS

(See Appendix)

ABBARNO, John, D'Youville College, New York	VI-2
ADAMS, Robert Merrihew, University of California/Los Angeles	I-3, VI-6
ALCOFF, Linda, Kalamazoo College, Michigan	VIII-1
ALLEN, Jeffner, State University of New York/Binghamton	X-5
ALLEN III, Paul, East Stroudsburg University	II-6
ANTON, Anatole, San Francisco State University	XI-2
ARCILLA, Rene V., University of Chicago	VI-1
ARNETT, Willard, Chatham College, Pennsylvania	VI-4
ATKINSON, Ti-Grace, Columbia University	II-2
ATWELL, John, Temple University	I-1
AUSTIN, Scott, Texas A&M University	V-4
BABICH, Babette B., Marquette University	II-4
BAER, Joachim, University of North Carolina/Greensboro	I-1
BALSKEV, Anindita Niyogi, Denmark	V-7
BARAD, Judith, Indiana State University	X-7
BARBER, Michael D., St. Louis University	IX-5
BARKER, Evelyn, University of Maryland	V-3
BAR ON, Bat-Ami, State University of New York/Oswego	XI-1
BARRAL, Mary Rose, Gannon University	II-1, V-3
BEAVERS, Anthony F., Marquette University	X-14
BELL, Linda Ann, Georgia State University	VII-1
BENDITT, Theodore M., University of Alabama/Birmingham	IV-2
BERGOFFEN, Debra, George Mason University	V-12
BERMAN, Robert, Xavier University	X-4
BERNAUER, James, Boston College, Massachusetts	X-3
BLACKWELDER, Brent, Environmental Policy Institute	X-9
BLAIR, Anthony, University of Windsor	VII-2
BOCK, Robert L., Western New England College, Massachusetts	III-6
BOOTHBY, Richard, Loyola College, Maryland	V-12
BRAND, Myles, Ohio State University	IV-2
BUFORD, Thomas, Furman University	V-9
BUTLER, Nance Cunningham, University of Connecticut Medical School	ol VI_9 Y_9
CAHOONE, Lawrence, Boston University	VII-3
CAMPBELL, Jean A., Brooklyn, New York	X-9
CARTWRIGHT, David, University of Wisconsin/Whitewater	I-1, VI-3
CATALANO, Joseph S., Kean College, New Jersey	VII-1
CAUMAN, Leigh, Columbia University	II-2
CAWS, Peter, George Washington University	III-5
CHAKREBARTI, Arindam, Oxford University	X-4
CHEN, Ellen May, St. John's University	III-4
CHENG, Chung-ying, University of Hawaii/Manoa	X-10
CLARKE, Norris, Fordham University	I-3
CONWAY, Daniel, Harvard University	II-4
CORRINGTON, Robert S., Pennsylvania State University	III-6, X-13
CUA, Antonio S., Catholic University of America	V-7
DANNHAUSER, Werner, Cornell University	
DANTO, Arthur C., Columbia University	IX-2 VI-4
DEES, J. Gregory, Yale University	IX-1
	11/1

APA PROGRAM, APPENDIX	601
Digitized by Arya Samaj Foundation Chennai and eGangotri	
DEJNOZKA, Jan, U.S. Naval Academy, Maryland	V-10
DEN UYL, Douglas J., Bellarmine College, Kentucky	IV-3
DESAN, Wilfrid, Georgetown University	III-5
DIFILIPPO, Terry, New York	V-1
DOMBROWSKI, Daniel A., Creighton University	X-7
DONNELLY, John, University of San Diego	IX-3
DORAN, Kathryn, Wellesley College, Massachusetts	V-11
DREYFUS, Hubert L., University of California/Berkeley	X-12
DUMAIN, Ralph	X-9
DUNLOP, Julie, Yale University	X-7
DURBIN, Paul T., University of Delaware	II-3
DURFEE, Harold A., American University	III-6
EMILSSON, Eyjólfur Kjalar, University of Iceland	VI-10
EMMANUEL, Steven M., University of California/Riverside	IX-3
ENLOE, Cynthia, Clark University	X-5
FARQUHAR, Dion, New York University	V-2
FELDMAN, Seymour, Rutgers University/New Brunswick	X-15
FERRÉ, Frederick, University of Georgia	II-3
FLYNN, Thomas R., Emory University	III-5
FREDE, Michael, Princeton University	VI-10
FRIEDMAN, Marilyn, Bowling Green State University	IX-4
GELBER, Sidney, State University of New York/Stony Brook	VII-3
GERAGHTY, Richard, Providence College, Rhode Island	II-8
GILLESPIE, Fran, President of CARA	I-4
GODFREY, Joseph J., St. Joseph's University	IX-5
GOTTLIEB, Robert S., Worcester Polytechnic Institute, Massachusetts	III-2
GRACIA, Jorge, State University of New York/Buffalo	X-15
GRAHAM, John T., University of Missouri/Kansas City	X-11
GROOTENDORST, Rob, University of Amsterdam, Holland	VII-2
GUNTER, Pete A. Y., North Texas State University	X-8
HABER, Honi, Temple University	VI-3
HALLETT, Garth, St. Louis University	V-11
HALPER, Edward, University of Georgia	X-4
HAMPSCH, George, College of the Holy Cross, Massachusetts	II-6
HANNON, Barbara, University of Arizona	X-6
MARMAN, Gilbert, Princeton University	II-7
HAROUTUNIAN-GORDON, Sophie, University of Chicago	VI-1
HATCHER, Donald, Baker University	II-7, X-12
INTIKKA, Jaakko, Florida State University	II-7
Tile, Shere, Hite Research	V-2
HOFFMAN, W. Michael Bentley College Massachusetts	IX-1
LIZBERGER, William G. Bucknell University	VI-4
CACK, Michael University of Maryland	IV-2
CLUAIL, Stephen DePaul University	X-4
LI, Kichard University of Wyoming	X-13
California Ctata University / Fullerton	V-7
X-V-III IIII Donnaulriania Ctata University	II-7
	X-14
Oly, David k II C Novel Academy Maryland	V-10
T. F. GWard I Introventer of Nove Orleans	X-14
KEENAN, James E. Fordham University	TT 1

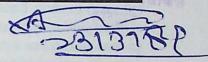
Digitized by Arya Samaj Foundation Chennai and eGangotri	
KELLY, David, New York	V-5
KEMP, Kenneth, U.S. Air Force Academy, Colorado	VI-7
VIM Lagreon Brown University	V-5
KLUBACK, William, City University of New York/Kingsborough Commu	inity
College	X-13
KOHAK, Erazim, Boston University	V-9
KONYNDYK, Kenneth, Calvin College, Michigan	I-3
KRONEGGER, Marlies, Michigan State University	II-5, X-10
KRUKS, Sonia, New School for Social Research, New York	III-5
KUNTZ, Paul G., Emory University	VI-4, X-8
LACKEY, Douglas, City University of New York/Baruch College	VI-7
LADD, Rosalind Ekman, Wheaton College, Massachusetts	V-13
LANGLEY, Raymond J., Manhattanville College, New York	II-8, III-6
LANGSDORF, Lenore, University of Texas/Arlington	VII-2
LARSON, Sue, Barnard College/Columbia University	II-2
LASKEY, Dallas, Concordia University	II-5
LAVELY, John, Massachusetts	V-9
LEE, Kwang-Sae, Kent State University	III-4
	2, X-3, XI-1
LEHMANN, Scott, University of Connecticut/Storrs	IX-4
LIMPER, Peter, Christian Brothers College, Tennessee	II-3
LINDEMAN, Kate, Mount St. Mary's College, California	IX-5
LOMASKY, Loren E., University of Minnesota/Duluth	IV-3
LONG, James R., Fairfield University	X-15
LOUGHNEY, John A., Westfield State College, Massachusetts	X-2
MACHAN, Tibor R., Auburn University	IV-3
MACK, Eric, Tulane University	IV-3
MAGNELL, Thomas, Smith College, Massachusetts	XI-1
MAGNUS, Bernd, University of California/Riverside	IX-2
MAHONEY, Edward P., Duke University	VI-8
MARKS, Joel, University of New Haven	
MARTIN, Bill, University of Kansas	V-7
MARTIN, Glen, Radford University	VI-5
MATTHEWS, Gareth, University of Massachusetts/Amherst	V-7
MCBRIDE, William, Purdue University	V-13
MCCARTY, Richard, East Carolina University	III-5, VI-2
MCCLENNEN, Edward F., Washington University	VI-3
MCLANE, Earl, College of William and Mary, Virginia	IX-1
MCLAUGHLIN, Andrew, City University of New York/Lehman College	IX-3
MELNYK, Andrew, Corpus Christic College, Oxford University, England	VIII-1
MEYERS, Diana T., University of Connecticut/Storrs	X-6
MILLER, Harlan B. Virginia Polytock 1. 1	IX-4
MILLER, Harlan B., Virginia Polytechnic Institute and State University MOHANTY, Jitendranath, Temple University	X-7
MORRIS Christopher W. Bowling C.	V-7
MORRIS, Christopher W., Bowling Green State University MORRIS, Phyllis S. University of Michigan State University	IV-3
MORRIS, Phyllis S., University of Michigan/Ann Arbor MORRIS, T. F., American University	VII-1
MORRISON, Donald, Rice University	IX-3
MOTHERSILL Mary Raymond C. II.	V-4
MOTHERSILL, Mary, Barnard College/Columbia University	II-2
MOTZKIN, Aryeh L., Boston University MUOIO, Patricia A., Maryland	VI-8
MURPHY Julien S. University CO.	X-9
MURPHY, Julien S., University of Southern Maine	VII-1

Digitized by Arya Samaj Foundation Chennai and eGangotri	
NEELEY, Steven, University of Cincinnati	VI-
NEHAMAS, Alexander, University of Pennsylvania	II-4, IX-5
NEVILLE, Robert C., Boston University	IV-1, X-2
NEWTON, Lisa H., Fairfield University	III-1
NICHOLSON, Carol, Rider College, New Jersey	VI-1
NORMORE, Calvin, University of Toronto & Ohio State University	VI-8
O'BARR, Jean F., Duke University	X-5
OLIVER, Kelly, Miami University	VIII-1
OLSON, Alan, Boston University	X-13
PARK, Sung Bae, State University of New York/Stony Brook	III-4
PARK, Ynhui, Simmons College, Massachusetts	V-3
PARKES, Graham, University of Hawaii/Manoa	V-7
PARSONS, Howard, University of Bridgeport	II-6
PEPPER, George, Iona College, New York	X-13
PINKARD, Terry, Georgetown University	IV-1
PITT, Joseph, Virginia Polytechnic Institute and State University	II-3
PLUHAR, Evelyn B., Pennsylvania State University/Fayette	X-7
PREUS, Anthony, State University of New York/Binghamton	V-4
PRIGOGINE, Ilya, University of Brussels & University of Texas	X-8
RASMUSSEN, Douglas B., St. John's University	IV-3
RAUCH, Leo, Babson College, Massachusetts	II-5
RECK, Andrew J., Tulane University	V-6
RESCHER, Nicholas, University of Pittsburgh	V-11
REUMAN, Robert, Colby College, Maine	VI-7
ROBINSON, Hoke, Memphis State University	X-1
ROHATYN, Dennis, University of San Diego	V-11
RORTY, Richard, University of Virginia	VI-1, IX-2
ROSE, Mary Carman, Goucher College, Maryland	I-4
ROSENBERG, Alan, City University of New York/Queens College	III-2, X-3
ROSENBERG, Jay F., University of North Carolina/Chapel Hill	IV-2
KOTH, Robin Alice, University of Scranton	II-4
RUSSOW, Lilly-Marlene, Purdue University	X-7
RYDER, John, State University of New York/Cortland	V-1
SAATKAMP, JR., Herman J., Texas A&M University	VI-4
JANIONI, Ronald E., Dennison University	VI-7
SAWICKI, Jana, University of Maine/Orono	II-2, V-2
OCHACHT, Richard, University of Illinois/Urbana-Champaign	II-4, IX-2
Olikag, Oswald () Fisk University	X-13
Oli I. Ofelia University of Florida	X-11
TUCK. IONI	X-5
SHAPIRO, Gary, University of Kansas	II-4, VII-3
TO LATER HELLE Visites I Injurgentity of Court Hounds	VI-7
AN Richard Temple University	II-4
Tarvey I minemite of Minemi	X-12
	V-12
The Inomes II cm	VI-5
	X-5
SINGER, Linda, Miami University SMITH, Barry, M.	V-2
SMITH, Barry, University of Manchester, England SMITH, Jadwigg Paid	X-6
SMITH, Jadwiga, Bridgewater State College, Massachusetts SMITH, John F. Vale University	X-10
John F Vala University	V O

SMYTH, Richard A., University of North Carolina/Chapel Hill	V-6
SNOW, James, Temple University	VI-3
SOLL, Ivan, University of Wisconsin/Madison	I-1
SORENSON, Roy, New York University	II-7
STACK George L. State University of New York/Brockport	XI-1
STAMBAUGH, Joan, City University of New York/Hunter College	V-7
STANTON, Betty, MIT Press	VI-4
STRANGE, Steven K., University of Pittsburgh	VI-10
STURM, Fred G., University of New Mexico	X-2
SULLIVAN, Dan, New Jersey	. X-14
THOMAS, Laurence, Oberlin College, Ohio	III-2
TONG, Lik Kuen, Fairfield University	III-4
UMPHREY, Stewart, St. John's College, Maryland	V-10
VAN EEMEREN, Frans H., University of Amsterdam, Holland	VII-2
VAN HEERDEN, Gita S., University of Massachusetts/Amherst	III-6
VER EECKE, Wilfried, Georgetown University	V-12
VON DER LUFT, Eric, Syracuse University	I-1
VOS, Arvin, Western Kentucky University	I-3
WALLACE, Kathleen, Hofstra University	VII-3
WALSH, George V., Salisbury State University	V-5
WALSH, Joe, Stockton State College, New Jersey	XI-2
WALTERS, Gregory, St. Mary's University	X-13
WARREN, Karen J., Macalaster College, Minnesota	V-13
WENISCH, Fritz, University of Rhode Island	I-4
WESTON, Anthony, State University of New York/Stony Brook	IX-4
WESTRA, Laura, University of Toledo	X-7
WILSON, Margaret, Princeton University	VI-6
WINFIELD, Richard Dien, University of Georgia	IV-1
WINKLER, Kenneth P., Wellesley College, Massachusetts	V-8
WISDO, David, Yale University	IX-3
WONG, David, Brandeis University	V-7
ZAMOSHKIN, Yuri, Institute of USA and Canada Studies/Moscow	11-6
ZOELLER, Guenter, University of Iowa	X-1
7WEIC Amulf II-in-	

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 11, NOVEMBER 1988



WHAT IS STILL VALUABLE IN HUSSERL'S ANALYSES OF INNER TIME-CONSCIOUSNESS*

NE of Edmund Husserl's enduring insights is that the temporality of consciousness is different from the time experienced in the objects of consciousness. In this paper, I shall examine Husserl's conception of the temporality of conscious life to determine how much of his theory can still be retained once one accepts that there is a real time and that there is other evidence about the temporality of consciousness besides first-person reflective evidence. I shall argue that Husserl's two-level theory of external time-consciousness and of inner time-consciousness is still the best available, but that it needs some important revisions.

Husserl's theory of time-consciousness is based on phenomenological evidence, but it is a theory. It is a theory that postulates that certain structures of consciousness explain how we are able to experience what we do experience. Phenomenological reflection provides the experiential data that the theory must explain. This experienced data is of three sorts: the experienced temporal features of the objects of consciousness, the experienced temporal features of consciousness itself, and any experienced temporal relations between consciousness and its objects.

The three main experienced temporal features of the objects of consciousness are temporal extension, temporal division, and temporal passage. These temporal features occur in the objects of many

* To be presented in an APA symposium on What Is Still Valuable in Husserl's Analyses of Inner Time-consciousness, December 30, 1988. J. Claude Evans will comment; see this JOURNAL, this issue, 617/8.

Zur Phänomenologie des Inneren Zeitbewusstseins (1893-1917), Rudolf Boehm, hrsg. (The Hague: Nijhoff, 1966). The Phenomenology of Internal Time-Consciousness, James Churchill, trans. (Bloomington: Indiana UP, 1966) is a translation of some of the material in the German volume. All section and page references in the text are to this book, first to the German and second to the English.

0022-362X/88/8511/0605\$01.20

© 1988 The Journal of Philosophy, Inc.

Digitized by Arya Samaj Foundation Chennai and eGangotri different types of mental acts, but they are most obviously experienced in perception. Temporal extension is being "spread out" through time. Physical objects and events are perceived to be spread out in time. To express this in another way, physical objects and events are perceived to have temporal parts which are in some way distinct from each other but which form a continuous whole. That temporally extended entities have somewhat distinct temporal parts (phases) is what I mean by temporal division. We perceive physical objects and events to be in some way divided into somewhat distinct phases that form a continuous series. Temporal passage is the transition from being future to being present to being past. Phases of physical objects and events are perceived to change at least from being present to being past, as in the case of perceiving motion.

Our own conscious life is also experienced to be temporally extended, temporally divided, and temporally passing. Although there are ontological differences between the temporality of consciousness and the temporality of worldly objects, the general experienced temporal features are similar, though not exactly the same. Perceptual acts, emotions, intentional actings, and other varieties of consciousness are all experienced as spread out through time and so as in some way divided into somewhat distinct phases that form a continuous series (at least for short periods). Conscious life is experienced as passing in that some phase is experienced as occurring now, while other phases are experienced as having just occurred, and still other phases are experienced as about to occur. Finally, for some varieties of consciousness, there are experienced temporal relations between consciousness and its objects. In perception and action, the present phase of consciousness is experienced to be simultaneous with the present phase of its intentional objects, and the past phases of consciousness are experienced to be simultaneous with the past phases of their intentional objects. The situation concerning the future is more complex.

Theories of external time-perception start from the idea that perceptual acts are themselves temporally extended and so are in some way divided into phases that occur at different times. For us to perceive at any given time the temporal extension of entities, the perceptual act-phase at that time must somehow grasp phases of the perceived entities that do not occur at that time. One of the traditional problems is how a perceptual act-phase is able to perceive those temporal parts of an entity which are not simultaneous with the perceptual act-phase. Formulated in A-series terms, how can consciousness perceive in the present those temporal parts of a temporally extended entity which do not happen in the present?

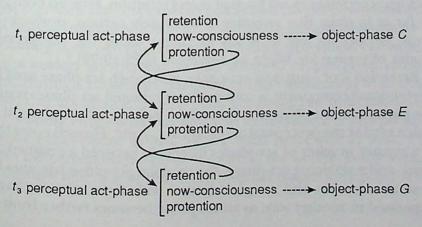
Digitized by Arya Samaj Foundation Chennai and eGangotri

The perception of the temporal passage of entities is inherently more complex than the perception of temporal extension. To perceive an entity as moving through time, we have to grasp not only phases of the entity which are earlier (or later) than the current phase, but also that those earlier phases were present and were perceived to be present. That the earlier phases were present but are no longer present (and that later phases are not present but will be present) is essential to the passage of time. That all of this is in some way perceived is what composes the perception of temporal passage. A mere succession of independent perceivings is not equivalent to the perceiving of temporal passage. We must somehow perceive the succession of perceptual act—phases along with the succession of worldly entities and the "transient presence" of both of these. Different theories of time-perception attempt to explain these phenomenological features through different structures of perception.

HUSSERL'S THEORY

Abstracting from the differences between two-level and three-level accounts, Husserl's answer to the traditional problems of time-perception is in his notion of the three features of any perceptual act. A perceptual act—phase has one feature that retains earlier phases of the perceptual act, another feature that perceives whatever is present, and a third feature that protends later phases of the perceptual act (see Diagram 1). It is through retention's direct contact with earlier perceiving and (more problematically) protention's direct contact with later perceiving that at any given moment perception of temporally extended entities can occur.

Diagram 1



Husserl contends that any perceptual act-phase intuits an object that has duration, that is, a temporally extended portion of the

Digitized by Arya Samaj Foundation Chennai and eGangotri

world. At any time, t_n , one feature of the perceptual act-phase perceives what is "now," that is, the object-phase that occurs at t_n . The retentional feature "reaches through time" to contact just-past phases which themselves have a now-consciousness feature that perceives an object-phase given as "now" (at t_{n-1} , t_{n-2} , etc.). Retention grasps the just-past phases as just-past and "looks through" them to their intended objects that are also grasped as just-past. Thus, retention's direct contact with the previous now-consciousness feature is also a direct contact with the object-phase perceived by this feature. Hence, to the presently perceived object-phase there is attached a continuum of just-past perceived object-phases that are experienced as just-past. This continuum of perceived object-phases stretching from the present moment backwards is the temporally extended object as perceived. In the example of the series of notes, we would hear the sequence C - E - G, even though precisely speaking the G alone was present, the E was just-past, and the C was further past.

The notion of retention is the crucial feature in Husserl's solution to the traditional problem of time-consciousness.

For only in primary remembrance do we see what is past; only in it is the past constituted, i.e., not in a representative but in a presentative way. The just-having-been, the before in contrast to the now, can be seen directly only in primary remembrance (§17:41;64).

Although retention is actual at one time, its intentional object is at an earlier time. Retention reaches to earlier moments in time and directly intuits earlier moments as earlier. As indicated in the above quotation, it is the direct givenness of the past as past. As the unmediated intuition of the past, retention has the same degree of epistemological certainty as any form of direct intuition. Husserl even goes so far as to claim that retention is absolutely certain, probably because he sees no way in which its claims about the just-previous phases of mental life might be corrected (§22:49;72).²

Retention is of a just-past act-phase. Since each act-phase itself has a retentional component, a continuum of retained act-phases results; the perceptual act-phase at t_n retains the perceptual act-phase at t_{n-1} , which retains the perceptual act-phase at t_{n-2} , etc. Husserl notes that, as a greater number of act-phases intervene between a retained act-phase and a retaining act-phase, the retained act-phase becomes less clear to retention and merges together with those surrounding it. Expressed in another way, as an act-phase becomes farther from the

² See also *Ideas*, W. R. Boyce Gibson, trans. (New York: Collier, 1962), §78, p. 203.

609

present, "the greater the blending and drawing together" ($\S9:26;47$) of it for present retention until it becomes completely imperceptible ($\S11:30/1;52$).

Husserl has very little to say about protention. In a number of places he portrays it as just like retention except that it concerns the just-future rather than the just-past (§26:55/6;79).³ As such, it would directly intuit later phases of mental life as just-future. Presumably his thought is that we have a direct intuition of the continuation of our mental life into the just-future. Husserl also says, however, that protention portrays the future emptily (§24:52;76; and §26:56;79-80)⁴, and that it portrays the future as indeterminate and open (Appendix 3:106;140; and §24:53;76).

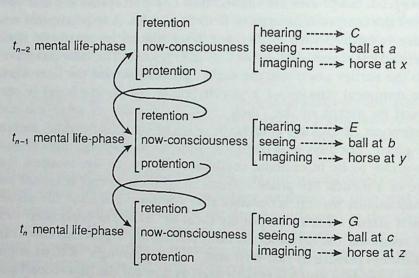
Husserl's theory provides a particularly insightful account of how perceptual act-phases are unified into one perceptual act that perceives the temporal passage of worldly entities. Any perceptual actphase is unified with earlier and later phases by means of their interlocking retentional-protentional structure. The later phases retain the earlier ones, and the earlier phases protend the later ones. The temporal passage of a worldly entity (or entity-phase) is perceived by means of the "look through" feature of retention and protention. Retention is directly aware of the just-previous mental life-phase, portrays it as just-past, and "looks through" this mental life-phase to its intentional correlates. The now-consciousness feature of a mental life-phase consists of all those components (of the simultaneous mental act-phases that compose a mental life-phase) which intend what is temporally focal or "now" as contrasted with the temporal background. In both portraying a previous now-consciousness feature as just-past and "looking through" it to its intentional correlate which is experienced as now for the now-consciousness feature, retention is able to experience a worldly entity-phase as just having been perceived as now. Similarly, protention can experience a worldly entity-phase as just about to be perceived as now. Thus, through its retentional-protentional unification with other perceptual act-phases, a perceptual act-phase can both perceive a later worldly entity-phase as now and experience the just-previous worldly entity-phase as just having been perceived as now. This allows the perceptual act-phase to portray the worldly entity as passing through time: the worldly entity's earlier phases were present but

³ Cf. also *Ideas*, §77, p. 198 and §81, p. 218. ⁴ Cf. also *Experience and Judgment*, James Churchill and Karl Ameriks, trans. (Evanston: Northwestern, 1973), §23b, p. 111.

are now past, its current phase is present but will be past, and its later phases are future but will be present.

I have discussed retention and protention so far only in the context of Husserl's response to the problem of "external" time-consciousness. In addition to our awareness of intended objects, we always have a nonreflective awareness of the variety of our conscious orientation. Although our focus is upon the intended object, there is always a concomitant awareness of perceiving it or remembering it or actively imagining it or some other mental act. Since these mental acts are extended through time, our experience of them must itself be a type of time-consciousness.

Diagram 2



A two-level analysis explains this inner time-consciousness by attributing a 3-feature structure like that of perceptual acts to all phases of mental life. All temporally extended mental acts have phases that occur at different times. Since several distinct mental acts can occur simultaneously, many mental act-phases can occur simultaneously. The set of mental act-phases occurring simultaneously (in one consciousness) is a mental life-phase. Each mental life-phase has a retentional, now-consciousness, and protentional feature. Diagram 2 represents this structure as applied to three concurrent mental acts.

Since the retentional feature intuits the entire previous mental life-phase (as well as its objects), there is an awareness through retention of the types of mental acts in which we have been engaged. Since the protentional feature intuits the entire just-future mental

life-phase (as well as its objects), there is an awareness through protention of the types of mental acts in which we will be engaged. There is even an awareness of the types of mental acts in which we are engaged at this moment, through the retention of the earlier mental life-phase's protention of the current mental life-phase. Thus, through the interlocking retentional-protentional structure, there is an awareness both of the types of mental acts and of their temporal extensions and simultaneity relations.

In the manuscripts assembled in Zur Phänomenologie des inneren Zeitbewusstseins, Husserl moved from a two-level account to a three-level account in which the third level is that of the absolute self-constituting flux which also is aware of immanent time (§s34-36, Appendix VI, and text 54). The version of Husserl's theory which I am defending is a two-level theory in which mental life occurs in a real time. I reject the third level because (1) there is no phenomenological evidence for it, (2) there is no theoretical reason to introduce it, and (3) its nontemporal or "quasi-temporal" (§s35/6, and Appendix VI) nature cannot provide for the awareness of a mental life that has A-series temporal features.

REVISIONS IN THE THEORY OF TIME-CONSCIOUSNESS

The basic Husserlian insights that time-consciousness depends upon the 3-feature structure and that retention "looks through" earlier mental life-phases that it grasps as past are sound, but the theory nevertheless needs some refinements and modifications.

- (1) The phases of mental acts, mental life, and worldly entities are conceived to be instants or infinitesimals, that is, to have a duration approaching zero. Whether or not this is the right idea for a real time, time as experienced should not be conceptualized in terms of infinitesimals.
- (a) Concerning mental acts and mental life, it should first be noted that infinitesimal phases are not observed in phenomenological reflection. What is observed is a continuity through time of mental entities. Infinitesimal phases are introduced for theoretical reasons to account for this experienced continuity through time (§16).

There are strong theoretical reasons, however, to deny that we experience infinitesimal phases of mental life. Infinitesimal temporal parts of mental life would introduce an infinite complexity into any mental life-phase. If there were infinitesimal phases, the retention of even one second of our past mental life would involve the retention of an infinite number of previous infinitesimal mental life-phases. Since the experienced continuity of mental life can be explained in another way, the experience of past durations of mental life should not be made dependent upon an infinite capacity.

The continuity of mental life means that there are no gaps in it. Mental life can be continuous even though the phases that are unified by the 3-feature structure endure unchanged through periods of real time. Such enduring mental life-phases retain and protend other temporally extended mental life-phases. Although the duration of such an extended 3-featured phase could be divided into smaller phases, these would not retain and protend each other. Rather, each smaller phase would participate in the retention of the same temporally extended phase that precedes the larger phase of which it is a portion. Thus, there would be a continuous mental life composed of temporally extended 3-featured phases that retain and protend each other.

Although mental life might fail to be continuous, the experienced continuity of mental life always exists in a consciousness that is unified by an interlocking retentional-protentional structure. Experienced continuity means that there are no *experienced* gaps in mental life. It requires that retention always find a mental life-phase at any moment of time that it contacts. Since retention by its very nature is of a previous mental life-phase, this condition is always satisfied, so it is satisfied by mental life-phases that endure unchanged through periods of real time. Individual mental acts, however, could be experienced to be discontinuous, because retention could contact a mental life-phase from which the act-phase was missing.

Experienced continuity could exist even if there were gaps in mental life in real time. So long as the gap was undetected, as in some cases of short "blackouts," there would be an experience of continuity.

(b) Concerning experienced worldly entities, we perceive their temporal extension by perceiving all of the entity-phases that compose the extension. If perceived entity-phases were infinitesimals, we would have to perceive an infinite number of them to perceive an entity to exist for one second. There is no good reason to place such a burden on perceiving temporal extension. The only advantage of infinitesimal phases is that they would make instantaneous changes in perceived worldly entities possible. Since the scientific data shows that we do not in fact perceive such instantaneous changes in content, however, there is no reason to make it possible for them to be perceived.

Perceived worldly entity-phases endure through real time but are not perceived as enduring, because there are no smaller perceived entity-phases that compose them. Although the enduring entity-phases can in principle be subdivided, these subdivided phases are

not distinguished in perception. The perceived continuity of entities through time requires only that there be a distinct worldly entity-phase that is perceived by the now-consciousness feature of every perceptual act-phase. Since perceptual act-phases also endure through real time, this condition is easily satisfied.

(2) Husserl thinks of retention as being infallible because it contacts just-past phases of mental life, but in fact retentions can be found to be mistaken. Retention can be inaccurate by portraying as having occurred experiences and actions that did not in fact occur. When we are engaged in many mental acts at one time, we may later (even moments later) retain an experience or action which we were expecting to happen but which never did happen. This sometimes happens when we have regularly perceived some type of occurrence, and we are busy at the time of today's token of that occurrence. Although we do not misperceive today's token when it happens, we do not pay any attention to it. Even immediately afterwards, we may retain the perception of the token with all of its familiar properties, even those which did not occur. For example, suppose that Mary always wears a blue suit to work. I see her today wearing a gray dress, but I take no particular notice of this because I am very busy with other activities. Even immediately afterwards, I may think and act on the basis of having just seen Mary wearing a blue suit. A similar sort of error in retention can occur when we retain having done some action which we intended to do but in fact did not do.

Retention is awareness of previous mental life-phases which are grasped as past and through which retention "looks" to their intentional correlates. In order both to retain a phase of mental life and to misportray some intentional correlate of that mental life-phase, we must fail in some way either in contacting or in "looking through" the mental act-phases of the phase of mental life. Although we could not retain the mental life-phase at all unless we successfully contacted and "looked through" some of its components, it does not seem to be necessary that we successfully contact and "look through" all of its components. We can contact a phase of mental life even though we are not aware of all of its acts and their intentional correlates.

In the above example, this would account for our not retaining Mary's gray dress, but it would not explain our sense that she was wearing the customary blue suit. This latter bit of information is not retained in the strict sense but added on in some other way. The possibility of such additions which are not experientially different from legitimate retentions means that retention (as determined ex-

perientially) cannot provide epistemological certainty concerning all of its apparent content. Nevertheless, the retentional feature of experience does provide extremely strong and reliable evidence.

(3) Husserl conceives the retentional-protentional unification of mental life-phases to operate without interests (to be a passive synthesis). Retentional awareness, however, does sometimes seem to be affected by our current focus of attention. When previously perceived data are irrelevant to that to which we are presently attending, they do not usually inform our present experience in the way that they would if they were relevant. For example, if we are listening to a recorded symphony or thinking through a theoretical problem and an important message interrupts this, the earlier moments of the symphony or steps of the theoretical problem do not significantly inform our attention to the message and the situation that it concerns. If our attention then shifts back to our earlier concerns, however, we may pick up the symphony or the problem at the point where we left it, relying upon our retention of the earlier moments.

Retention of any mental life-phase other than the just-previous one operates through the just-previous one, which operates through the phase just-previous to it, etc. Because of this and the fact that we cannot change the past, we cannot retain an experience that occurred two minutes ago unless we have retained it all along. This means that, if at any time our present experience is informed by the retention of an earlier experience, we must have been retaining it throughout the entire period since its occurrence. The changes in retention in the type of cases mentioned above present a problem, because it appears that we resumed the retention of the experience after a period during which we were not retaining the experience.

Since there definitely is some experiential difference in this type of case, and it seems to be a difference in the retentional feature, some aspect of retention must account for it. The simplest and most direct explanation is to postulate two essential aspects of retention: (a) an automatic interestless "maintaining" or keeping available for consciousness (in its horizon), and (b) a degree of prominence or importance of what is retained to present experience, which can vary according to our present involvements and understandings of what is perceived. The maintaining aspect of retention would operate independently of interests and would be that which unifies phases of mental life. The maintaining would have all of the structures of retention as previously described. Whatever was maintained, however, could inform our experience to different extents ranging from having very little prominence to being an essential part of the mean-

ing of our current mental acts and their intentional objects. Some degree of prominence would have to exist if retention is to be a feature of conscious experience, but a retained mental life-phase could inform present experience to a minimal extent in so far as we have some background sense of where we are and what we had just been doing.

(4) Protention requires an analysis that is different from that for retention. There are insurmountable problems with the conception of protention as direct awareness of just-future mental life-phases.

Commonsensically, we think of the future as being open and at least partially indeterminate and the past as being fixed and fully determinate. On this version of an A-series time, protention cannot be just like retention, because there is not a fixed and fully determinate future mental life to be contacted.

Protention could be analyzed in the same way as retention if three conditions were satisfied: (a) mental life and worldly entities exist in a B-series time so that there is (tenselessly) at any given date a definite mental life-phase with its intentional correlates, (b) this B-series time allows some mental entities to exist across temporal locations, rather than to have to exist completely at a temporal location, and (c) protention has the same very high degree of accuracy as retention. Condition (a) would eliminate the asymmetry between the past and the future, and condition (b) would allow protention to "reach across" temporal locations. At any given time there would be a fixed and fully determinate future capable of being contacted by protention.

Condition (c) presents a problem on any ontology of time, however. Our protentions are far more fallible than our retentions, and it seems to be possible to be mistaken about all of the components of a future phase of mental life. Assuming that there were fixed and fully determinate future mental life-phases, it does not seem possible that a protention could contact such a future phase in the way a retention contacts past phases, yet still be mistaken about all of its components.

The fallibility of protentions remains a problem even if an A-series ontology of time is assumed. The most promising way to interpret protention as direct contact under an A-series time is to conceive it as contact with a set of alternatives. In an A-series time, the future-as-future can be conceived as a set of relatively determinate alternatives, only one of which will become present. Protention can then be contact with a set of alternatives with one singled out. This might appear to resolve the problem of fallibility, because only the singled

out phase of mental life fails to become present; one of the other alternative phases of mental life does become present, and protention is in contact with it (although as a non-singled-out alternative).

There are two problems with this purported solution. (i) There is no experiential evidence that protention grasps a set of alternatives. Our immediate anticipation of our just-future mental life and its correlates is of single courses of perception, thought, and action. The notion that protention really is aware of a set of alternatives does not have any phenomenological basis and seems to be simply ad hoc. (ii) Even if protention is allowed to be contact with a set of alternatives, there is no reason to think that we immediately anticipate the right set of alternatives. Events that are complete surprises to us do happen; it is hard to understand how this could be the case if we regularly contacted the real future alternatives.

(5) Because protention does not directly contact later phases of mental life and their intentional correlates, the unification of mental life-phases necessary for time-consciousness is accomplished primarily by retention. Only present and past phases of mental life are involved in the genuine perception of the temporal extension and the temporal passage of worldly entities. Of course, through protention there is an anticipation of the phases of mental life and their intentional correlates to come, and these are generally accurate. The depicted mental life-phases and their intentional correlates are, however, ontologically distinct from the phases that become present.

PETER K. McINERNEY

Oberlin College

Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENT ON MCINERNEY*

ETER McInerney argues that, according to Husserl, the retentive consciousness of elapsed temporal phases of perceived entities is a function of the retentive consciousness of earlier phases of the perceptual act itself. This retention "looks through" the retended act-phases to their intended objects. It is not clear that this is in fact Husserl's view, since it would seem to involve a revision rather than an interpretation of what Husserl calls the "double intentionality" of retention. One might find support for this position in the fact that Husserl's analyses of time-consciousness generally deal with the constitution of sensations as immanent objects rather than with the perceived objective tones themselves, but such an interpretation would make crucial use of precisely the weakest aspects of Husserl's phenomenology of perception: the dualism of the hyletic contents of sensation versus intentional form. From Aron Gurwitsch to Jean-Paul Sartre and Maurice Merleau-Ponty, this doctrine has been subjected to a devastating critique, and it would seem clear that the phenomenology of time-consciousness has to start with the basic structure of consciousness of the world, and not with the artificially isolated constitution of immanent objects.

Ι.

Once the dualistic theory of perception has been abandoned, we are in a better position to reappraise Husserl's theory of absolute timeconstituting consciousness as a theory of self-constitution. Husserl introduces the absolute stratum, constituting but not itself constituted, in order to account for the self-appearing of the stream of conscious life as a stream. McInerney's suggestion that Husserl's three-level account be replaced by a two-level account is, I think, important. Rather than posit a flowing but, strictly speaking, nontemporal (because not itself constituted as a temporal unity) consciousness of our intentional life, McInerney sketches the outlines of a genuine self-constitution of conscious life. The most interesting point here is that in Diagram 2 there is no direct consciousness of the now-consciousness (i.e., of the current phase of the perceptual act itself): the only consciousness of now-consciousness is its protention in a currently retended past phase of consciousness and its retention in a phase that is currently protended as retending the present phase.

0022-362X/88/8511/0617\$00.50

© 1988 The Journal of Philosophy, Inc.

^{*} Abstract of a paper to be presented in an APA symposium on What Is Still Valuable in Husserl's Analyses of Inner Time-consciousness, December 30, 1988, commenting on a paper by Peter K. McInerney, this JOURNAL, this issue, 605–616.

Digitized by Arya Samaj Foundation Chennai and eGangotri

There is, as it were, a "blind spot" at the very core of the reflexivity of consciousness. This analysis supports a nonegological theory of consciousness as developed by Gurwitsch. It remains to be seen whether it is also the opening for Derridean différance. I think that this result suggests that the model for the analysis of time-consciousness cannot be a theoretically oriented perceptual consciousness, as is the case in much of Husserl's work. The analysis must rather be oriented to more "actional" phenomena along the lines of what Husserl later calls "primal striving." Such an orientation removes the temptation to identify passive synthesis with the absence of interests.

TT

I think that it can be shown that McInerney reifies Husserl's "infinitesimal phases." Husserl makes it clear that momentary phases are "ideal limits," and that what is in fact experienced is always a certain "thickness" of time which is not composed of ideal limits as elements.

J. CLAUDE EVANS

Washington University

FEMINIST SKEPTICISM AND THE "MALENESS" OF PHILOSOPHY*

N the late 1970s in the United States, contemporary American feminism took an important turn. From an initial emphasis on legal, economic, and social discrimination against women, feminists began to consider the deep effects of the gender organization of human life on Western culture—on the literary, scientific, and philosophical canon that we call "the Western intellectual tradition." Earlier feminist works had criticized that tradition for its explicit gender biases: objectionable images of women, misogynist theory, the lack of representation of women's concerns and voices, etc. In the late 1970s, however, a deeper "hermeneutics of suspicion" emerged among feminists. We began to realize that gender bias may be revealed in one's perspective on the nature of reality, in one's style of thinking, in one's approach to problems—quite apart from any explicit gender content or attitudes toward the sexes.

Recently, some contemporary feminists have taken yet another critical turn. Criticizing what they see as the historical oversimplifications and unconscious ethnocentrisms of earlier feminist readings of culture, more recent perspectives urge a new caution, a new skepticism about the use of gender as an analytical category. These perspectives, like much contemporary thought, are informed by what might be called a "theoretics of heterogeneity"—an attunement to multiple interpretive possibilities, to the plurality of interpenetrating factors that comprise any object of analysis, and to the "differences" that fragment all general claims about culture. According to such perspectives, to theorize culture or history along gender lines—to speak of "male" and "female" realities or perspectives—is to homogenize diversity and obscure particularity.

Although most commonly associated with "postmodern" continental perspectives, such a "theoretics of heterogeneity," particularly with respect to its implications for a cultural understanding of philosophy, is strikingly and articulately exemplified by Anglo-American analyst Jean Grimshaw. In *Philosophy and Feminist Thinking*,

Philosophy and Feminist Thinking (Minneapolis: Minnesota UP, 1986).

^{*}To be presented in an APA symposium on Feminist Skepticism, December 30, 1988. Naomi Scheman will comment; see this JOURNAL, this issue, 630/1.

I first formulated several ideas developed here in an unpublished manuscript, co-authored with Alison Jaggar. I thank her for conversations and editorial suggestions from that collaboration. I also thank Jonathan Schonsheck, Lynne Tirrell, and especially Lynne Arnault.

emphasizing what she describes as the "extremely variegated nature" of human experience (ibid., p. 102), Grimshaw presents many skeptical challenges to the notion that the history of philosophy can meaningfully and nonreductively be characterized as "male." Although I share Grimshaw's pluralist sensibilities, my purpose in this paper nonetheless will be to argue for the legitimacy of the project she criticizes. There are complexities and cautions that must be taken into account in approaching the question of philosophy's "maleness." But, I argue, the question is neither ill-founded nor unanswerable. Due to space limitations, it is not possible to consider Grimshaw's arguments in all their variety and specificity. Instead, working with categories and insights derived largely from French feminism, I propose a general reading of the "maleness" of philosophy, and then interrogate that reading along the lines of Grimshaw's critical concerns. Although I defend my reading against Grimshaw's skepticism, I also indicate a context in which the insights of her skepticism may be profitably employed. Finally, I briefly discuss some of my own concerns about gender skepticism as a potential theoretical turn within feminism.

Ι.

Before I do any of these, however, it is necessary to present some theoretical background to the issue. In the United States, two works in the 1970s—Dorothy Dinnerstein's *The Mermaid and the Minotaur*² and Nancy Chodorow's *The Reproduction of Mothering*³—were especially influential in charting two important directions that feminist conceptions of the "maleness" of philosophy were to take.

Dinnerstein's central focus is on destructive cultural attitudes toward nature, the human body, mortality, and sexuality. These attitudes are gender-biased, Dinnerstein argues, not in the sense that they are exclusively or even distinctively held by men, but rather in the sense that they ultimately derive from a *gynophobia* to which both men and women are vulnerable, insofar as they have been raised within a system of female-dominated infant care. Within such a system, everything pleasurable and everything terrifying about bodily needs, desires, and vulnerabilities are first experienced in the arms of a woman—the mother. As a consequence, the entire arena of spontaneous bodily experience becomes associated with "woman" in general and remains split off from the cultural psyche—now

² New York: Harper & Row, 1977. ³ Berkeley: California UP, 1978.

identified with all that is "mind," and coded as male. Split off in this way, our ambivalence toward the body remains culturally unintegrated and destructive, particularly for those on whom it is "projected": woman and nature.

Dinnerstein's cultural emphasis has affinities with developments in French feminism. Inspired by Lacanian psychoanalysis, the key category here is that of "phallocentrism," the reign of the phallus operating as a metaphor for the cultural privileging of unity, stability, identity, and self-mastery over the "maternally connoted" values of body, spontaneity, multiplicity, loss of self, etc. Both men and women, Julia Kristeva argues, find reality constructed for them through the template of this symbolic system; for her (although not for other French feminists), there is no specifically feminine form of discourse. Women, insofar as they participate fully in the dominant symbolic order, are just as "phallic" as men; some men, on the other hand (e.g., iconoclastic artists and philosophers) have often been inspired disrupters of that order.

Chodorow's emphasis, in contrast, is on differences in the psychology and "cognitive style" of men and women, differences stemming from contrasting patterns in the development of male and female infants. Because a more rigorous individuation from the mother is demanded of boys, they grow up, Chodorow argues, insisting on clear and distinct boundaries between self and others, self and world, and defining achievement in terms of emotional detachment and autonomy.

Possibilities for the application of Chodorow's ideas to the analysis of our male-dominated intellectual traditions immediately presented themselves to feminists from many disciplines. Adopting her emphasis on gender difference, many American feminists began to see the individualist and objectivist biases of dominant traditions in the disciplines and professions as "masculine"—that is, as connected to characteristic features of the construction of male psychology and personality. Carol Gilligan's influential In A Different Voice⁵ emerged from this stream of feminist thought. In it, she elaborates on the consequences of the privileging of detachment and individual autonomy in developmental theory and in our dominant ideals of moral reasoning. Created by men, those theories and ideals establish as normative a view of social relations and a hierarchy of ethical

Julia Kristeva, About Chinese Women (New York: Urizen, 1977).
 Cambridge: Harvard, 1982.

values which, Gilligan argues, do not reflect the more relational picture of reality underlying the ethical reasoning of women.

The feminist critique of the Western philosophical tradition has been nourished by insights from both these arenas of feminist thought. Writing from a variety of perspectives, feminist philosophers have by now produced a formidable body of gender analysis directed at the traditions of Western philosophy and specific authors within it.⁶ Many argue that philosophy has been "masculine" in a psychological sense; others have focused on the persistence of gyno-

phobic themes or phallocentric structurings of reality.

Feminist controversy surrounds the historical generalizations involved in such characterizations of philosophy. Grimshaw, pointing to the distinctive individual perspectives and major conceptual transformations which have marked the history of philosophy, concludes that the traditions of Western philosophy are too heterogeneous to permit any generalizations, gendered-based or otherwise. Hélène Cixous⁷ and Luce Irigaray,⁸ in contrast, work with categories derived from both Simone de Beauvoir and Jacques Derrida, and argue that a distinctive form and a characteristic logic runs throughout the history of Western philosophy. In the next section, I shall address some issues involved in such a dispute.

TT

According to Cixous and Irigaray, the characteristic form of Western philosophy is that of the hierarchical opposition (or "dualism," as Anglo-American philosophers would call it): the bifurcation of reality into mutually opposed elements, one of which is privileged and identified with "self," the other of which is disdained and designated as "other." Although the specific content of this structuring is not fixed, indeed varies culturally, what remains constant is the binary logic of the metaphysics, which posits a unitary identity, a center, a positivity—and then, defined against that positivity, that which is different, negative, and marginalized. For example, we see a frequent philosophical identification of "self" with "mind," and "body" with threatening "other." Western philosophy's historical obsession with "pure" thought, its search for ultimate or foundational categories with which to order the world—these projects, too,

⁶ See the APA Newsletter on Feminism and Philosophy (October 1988) for a representative bibliography.

^{7 &}quot;The Laugh of the Medusa," Signs, 1 (Summer 1976): 875-893; and (with Catherina Clement) The Newly Born Woman (Minneapolis: Minnesota UP, 1986).

8 Speculum of the Other Woman (Ithaca: Cornell, 1985).

assume a hierarchical, oppositional construction of reality. For the hope that they may succeed depends upon the conceptualization and circumscription of a privileged human faculty—"reason"—whose essential relation to objects is transparent and pure (Richard Rorty's "mirror of nature"), and which is capable of transcending that which threatens to mire it in obscurity and chaos: emotion, perspective, human interest, and so forth. The agonistic struggle implied in this description derives from the precariousness of the dualist construction itself: on some level, the Other is felt to be the part of the self which it actually is; hence, the necessity for constant vigilance against it.

In proposing this structural reading of the history of Western philosophy, I am not denying that this history has been marked by radical conceptual change. Nor am I claiming that this structure dominates seamlessly, or without resistance. Many counterexamples can of course be produced, demonstrating the existence of perspectives that are not hierarchical, not concerned with purity or foundations, and so forth. Grimshaw frequently presents such examples (Nietzsche's perspective on the body, for example) as counterevidential against various feminist attempts to describe patterns and coherencies in the history of philosophy. Such a strategy, however, overlooks the fact that dominance does not require homogeneity in order to function as dominant. This point applies both within the philosophical conversations of particular eras and across them.⁹

Even if Grimshaw were to concede that French feminism's structural reading of the history of philosophy does provide at least some partial illumination of that history, the question would rightfully remain: In what sense, if any, is the dualist structuring of reality "male"? Clearly, dualism is not "male" in any innate sense—for many male philosophers have protested strongly against our dualist traditions. The question is, rather, whether those male philosophers who have dissented from dominant philosophical constructions of reality have, in doing so, deviated from constructions of masculinity dominant within philosophy.

Some insight is provided by the fact, pointed out by French and American feminists alike, that the hierarchical oppositions of Western thought have consistently been gender-coded. Characteristic of our philosophical traditions, for example, is the conceptualization of

 $^{^{9}}$ In a longer version of this paper I elaborate these claims, drawing on the ideas of Michel Foucault.

reason as a distinctively male capacity. By contrast, those faculties against which reason variously has been defined, and which it must transcend—the instincts, the emotions, sense perception, material-

ity, the body-typically have been coded as female. 10

If Dinnerstein and Kristeva are right, however, such coding is a consequence of "maternal connotations" and associations which both men and women develop through the infantile experience of body care by a woman. In that case, responsibility for such a symbolic system is appropriately attributed to the patriarchal division of labor. which has designated the woman as chief infant caretaker. This is not the same, however, as connecting hierarchical dualisms to a distinctively "masculine" perspective on reality. To establish that, one needs to show that there are significant differences in the perspectives of men and women, differences which dispose men, at least more so than women, to see the world dualistically.

One way of approaching this, as noted earlier, is through examination of differences in the psychological development of males and females—for example, the work of Chodorow. This approach, in its focus on patterns of child rearing specific to the bourgeois nuclear family, has obvious limitations. Chodorow does not acknowledge the historical specificity of the developmental scenarios she describes for example, an unprecedented emphasis on infant nurturing and on the socialization of children into "masculine" and "feminine" roles. But, clearly, such emphases, developed alongside industrialization and the rigid designation of "public" and "private" spheres, cannot provide illumination of the construction of male personality during the classical period of Greek philosophy or in the Middle Ages (when it is not even clear that childhood existed, let alone a gendered childhood). For this reason, analyses such as Chodorow's can have only a limited historical application to the interpretation of philosophical themes and concerns.11

The fact that one cannot provide an account of philosophy's "maleness" in terms of some transhistorical male personality development does not entail, however, that there are no transhistorical practices—child bearing, for example—that have conditioned West-

10 Cixous and Irigaray, op. cit.; and Genevieve Lloyd, The Man of Reason (Min-

neapolis: Minnesota UP, 1984).

⁹ In a longer version of this paper I elaborate these claims, drawing on the ideas of Michel Foucault.

Thus circumscribed, however, they may be tremendously useful—for example, in exploring the Western "Enlightenment" emphasis on objectivity, autonomy, and individual rights. It is striking that such an emphasis emerges historically alongside the development of the modern bourgeois family and its ideologies of autonomous, public "masculinity" and nurturant, Other-oriented "femininity."

ern men's and women's perspectives in significantly different ways. For Grimshaw, of course, the "lack of consensus" (as she calls it) in the interpretation and conceptualization of child bearing precludes the possibility of legitimate generalizations along such lines. Women have perceived child bearing, she notes, "as both the source of their greatest joy and as the root of their worst suffering" (op. cit., p. 73). The existence of multiple valuations of events such as giving birth ought not, however, obscure the fact that there are bodily realities involved in the praxis of reproduction—realities which are unavailable to men, and which one would have to be a very good Cartesian to dismiss as epistemologically irrelevant. The specific consequences of this for the ways in which space, time, embodiment, self, and other are constructed are not obvious, of course. But that there may well be a basis for a meaningful exploration of Western philosophy as gender-inflected ought not, certainly, to be dismissed a priori. 12

Even if Grimshaw were to allow that meaningful "gender elements" transversing cultural diversity exist (an assumption hotly debated among feminists), she would insist that we are methodologically prohibited from isolating and describing them:

The experience of gender, of being a man or a woman, inflects much if not all of people's lives . . . But even if one is always a man or a woman, one is never *just* a man or a woman. One is young or old, sick or healthy, married or unmarried, a parent or not a parent, employed or unemployed, middle class or working class, rich or poor, black or white, and so forth . . . Experience does not come neatly in segments, such that it is always possible to abstract what in one's experience is due to 'being a woman' from that which is due to 'being married,' 'being middle class' and so forth (op. cit., pp. 84/5).

Grimshaw questions here whether we can abstract the "gender component" from the heterogeneous constellations of factors that interpenetrate with it. I would agree with her—if what is at stake here is a Cartesian project, seeking to isolate and quantify "clear and distinct" elements from the objects under investigation. We shall, of course, never find that kind of clarity in the manifestation of gender, race, class—or any other variables, for that matter. But, as anyone who has taught courses in gender knows, there are many junctures at which, for example, women of color and white women discover pro-

¹² Cf. Iris Young, "Pregnant Embodiment: Subjectivity and Alienation," *The Journal of Medicine and Philosophy*, 1x (1984): 45-62, for a critique of philosophical dualism via its exclusion of experiences of embodiment such as pregnancy.

found commonalities in their experience (as well as profound differences.) One can, of course, adjust one's methodological tools so that these commonalities become indiscernable under the finely meshed grid of various "inflections" (or the numerous counterexamples that can always be produced.) When it comes to issues of race, however, intellectuals have no problem allowing that color (that is, the experience of being a person of color in a racist culture) creates some similarities of position across class and gender, and that these can be retrieved and discussed. What is not clear is why we have such a problem when it comes to issues of gender.

III.

Grimshaw presents her skeptical methodological point as an argument against the legitimacy of specifying "male" and "female" points of view. In the context of the history of philosophy, however, her point may be employed differently—to *illuminate* rather than dispute the "maleness" of philosophy. For a particular intersection of gender, race, and class has been typical of Western philosophy, and only in terms of such an intersection can we get a handle on the particular constructions of masculinity which have informed the history of the discipline.

The authors of our classical philosophical traditions have predominantly been white, economically and socially privileged males. We have no way of imagining what their work would look like if they had written "only" as white, or "only" as males. Nor can we isolate the "maleness" of philosophy from the context of dominance and subordination which has constructed gender, class, and race in our culture. The fact that philosophy has been dominated by white, privileged men has meant that it has developed from the center of power rather than from the margins of culture. This has surely had as significant an effect on its development as the fact that it has been authored by men. But, given the particular conjunction of gender and power characteristic of Western history, we have no way of separating these elements out. The "phallicness" of Western metaphysics is inseparable from its "centrism."

It should be no surprise when we find, then, that dominant philosophical dualisms are overdetermined with respect to race, gender, and class. So, for example, the hierarchical, oppositional construction of reason/unreason may be coded in a variety of ways, including (but not exclusively) as male/female. We also find "unreason" coded as African (e.g., in Hegel), and associated with practical, manual activity (a continual subtheme, as John Dewey has pointed out, in Greek philosophy). But despite this flexibility with regard to the

content of the duality, its hierarchical form is constant: it posits a revered identity (reason, mind, spirit, the white race, the male sex, and so forth), and then defines what is different, Other, and inferior to it (body, matter, practical activity, woman, African).

This structuring has served a multitude of philosophical and ideological purposes; it is overdetermined, after all, by the intersection of three separate axes of privilege and power. As regards gender, it can be seen that, through the consistent philosophical identification of women with the bodily arena of "unreason," a powerful ideological support is created for keeping women in their "material" place, excluded from those activities seen as requiring rationality and objectivity. Instead, women are relegated to the corruptible, "material" realm of which the "man of reason" wishes to have no part: the mundane care of the body, food preparation, the management of everyday dirt and disorder, and so forth.

Here, it is particularly and bitingly ironic when conceptions of philosophy employ metaphors of "housecleaning" in describing the task of philosophical analysis: Arthur C. Danto's 13 "conceptual housekeeping," David Raphael's14 "mental clearance." Raphael even reminds us, in the style of a Good Housekeeping article, that mental clearance, like cleaning the house, "is not a job that can be done once and for all. You have to do it every week. The mere business of living continues to produce more rubbish, which has to be cleared regularly" (ibid., p. 16). Raphael here picks up on Locke's famous description of the philosopher as an "under-labourer . . . removing some of the rubbish that lies in the way of knowledge."15 But such metaphors evoke the realm of the material, practical, and everyday precisely to offer a vision of their transcendence. The conceptual housekeeper "executes" his tasks, it turns out, in a hierarchial, dualistic neighborhood within which it is the Other that makes all the mess. For Danto, this messy side of town is inhabited by the other disciplines, constantly generating conceptual confusion, but too "robustly busy," as he says, to tend to the cleaning of their own houses. W. V. O. Quine's use of the cleaning metaphor is even more suggestive of class associations: he speaks, in Word and Object, of the philosopher's task as "clearing the ontological slums." 16

It is not women alone, then, who have been relegated to the "ma-

¹³ What Philosophy Is (New York: Harper & Row, 1968), p. 10.

¹⁴ Problems of Political Philosophy (London: Macmillan, 1976), p. 16.
15 Locke, "The Epistle to the Reader," in An Essay Concerning Human Understanding (New York: Dutton, 1961), p. xxxv.

terial" arena. But being overdetermined by gender, race, and class is not equivalent to being undetermined by gender. The hierarchical dualisms of Western culture did not descend from an androgynous heaven; they are the historical product of a white, Western male imagination of reality. It is difficult to conceive of any marginalized or subordinated group dreaming up such a metaphysics—although, of course, we often perpetuate and reproduce it, living within a culture that is organized, though not seamlessly, along such lines. Examples of women who "think like men" do not invalidate correlations of gender and perspective; rather, they prove the power of our male-created institutions and symbolic systems. At the same time, the position of marginality, as Kristeva argues, confers a certain potential for disruption of those systems. It is no accident, surely, that the beginnings of the current philosophical re-assessment of objectivism and foundationalism followed close on the heels of the public emergence, in the 1960s and 1970s, of those groups marginalized by the dominant metaphysics.

IV.

Feminist skepticism has operated as a necessary corrective to the sometimes unitary and universalizing notions of identity, perspective, and voice which emerged from early gender theory. That theory, usually based on the experiences of white, middle-class men and women, has often been guilty, as Grimshaw and others rightly have pointed out, of perpetuating the exclusion of difference which has been characteristic of "male-normative" theories. But, while we need to guard against facile and homogenizing generalizations about gender, we can also err, I believe, in the direction of multiplicity and difference. Let me conclude with some concerns along these lines.¹⁷

The feminist "en-gendering" of Western thought was instrumental in demystifying and intellectually dismantling the philosophical icon of an abstract, universal reason, a reason without race, class, gender, or history (the "view from nowhere," to borrow Thomas Nagel's phrase). There is no "view from nowhere," feminists insisted; all thought is socially located. Disquietingly, skeptical feminism may be in the process of installing a new version of the "view from nowhere," by deconstructing gender right out of operation as a

¹⁷ Due to space limitations, I can only touch briefly on these concerns and on the feminist critique of ethnocentrism in gender theory. I elaborate more fully in "Feminism, Postmodernism and Gender–Scepticism" in *The Politics of Method: Feminism and Postmodernism*, Linda Nicholson, ed. (New York: Routledge, Chapman & Hall, forthcoming).

¹⁸ The View from Nowhere (New York: Oxford, 1986).

tool for relocating reason from the Cartesian heavens of disembodied rationality and into the bodies of actual human beings. If philosophy cannot, in some meaningful sense, be said to reflect the gender of its creators, how can it be said to reflect their race, their class, or their historical place?

We need to take care that a "theoretics of heterogeneity" not lapse into a purely particularistic mapping of culture. Grimshaw's "inflection" argument, for example, although designed to display the fragmented nature of gender, in fact deconstructs race, class, and historical coherencies as well. For the inflections that modify experience are, of course, endless. (Race, class, and gender are the current "big three"; but what about religion? Age? Sexual orientation?) Some item of "difference" can always be produced which will shatter any proposed coherencies. If generalization is only permitted in the absence of multiple inflections or interpretive possibilities, then cultural generalizations of any sort-about race, about class, about historical eras—are ruled out. What remains is a universe composed entirely of counterexamples, in which the way men and women "see" the world is purely as particular individuals, shaped by the unique configurations that form that particularity. At this juncture, social critique self-destructs, and the dualistic, hierarchical nature of the actualities of power in Western culture is obscured.19

I believe that feminists must hold fast to the analytical category of gender and not allow it to become lost in a Sargasso Sea of counterexamples and endlessly multiple meanings. Feminist efforts have now begun to produce visible results in the curriculum, as more and more philosophers acknowledge patterns of exclusion in our discipline and reconceive their courses accordingly. But there are others who would reinstate the "view from nowhere" (read: the Great Works, undisturbed by gender, race, class, or cultural analysis) to the center of our curriculum, and rethrone the future philosopher-king as the only student worth teaching. At such a juncture, we may pay a very high institutional price for our intellectual deconstructions.

SUSAN BORDO

Le Moyne College

¹⁹ Lynne Arnault makes a similar point in "The Uncertain Future of Feminist Standpoint Epistemology" (unpublished manuscript).

FURTHER THOUGHTS ON A "THEORETICS OF HETEROGENEITY"*

USAN Bordo properly holds the feminist philosopher intellectually responsible both to the history of philosophy, which she (or he) attempts to illuminate critically, and to feminism as a political movement, which she aims to advance. Such dual loyalties are by no means new to philosophy: seventeenth-century philosophy, for example, is inseparable from the social, economic, and political transformations out of which it arose and which it served.

I want to discuss, therefore, not the propriety of combining these two aims, but an interesting intersection of concerns which illuminates them both. As Bordo makes clear, the question of differences is important both to the feminist critique of philosophy and to feminist theory. I shall suggest that the issue of differences among those who are culturally recognized as epistemic subjects provides both an analytic tool for a feminist critique of philosophy and a way of thinking about the politics of the women's movement and of other contemporary liberation struggles.

Heterogeneity as a philosophical problem goes back at least to the pre-Socratics, although probably only with the emergence of what we commonly recognize as fully *philosophy* did attention focus on the heterogeneity of the *knowers*, rather than on the *known*. In particular, Plato expresses a concern, which continues throughout the subsequent history of philosophy, with how language could hook onto the world, given the variable and contingent nature of our acquaintance with objects and the consequent contingency of the names we apply to them. Similarly, the variability in perspectives on, and opinions about, the nature of the world or about what is right or wrong has been taken to call into question our ability to know anything about what is or what ought to be.

There have been enormous changes in the philosophical responses to these concerns. What has been continuously hegemonic (not, as Bordo perceptively points out, uncontested) has been the attempt to render epistemic heterogeneity merely apparent. Epistemic authority has been granted to those who had some socially legitimated claim to transcending differences among individual knowers. What has changed historically are the terms of such authority. In particular,

^{*} Abstract of a paper to be presented in an APA symposium on Feminist Skepticism, December 30, 1988, commenting on a paper by Susan Bordo, this JOURNAL, this issue, 619-629.

the epistemology of modernity has granted authority to individuals in their own right, whereas in premodern Europe the claim to authority required tracing a line of legitimation up to the unitary authority of the king or God.

Modern epistemology places the legitimation that comes from a denial of fundamental differences inside the psyche of individual knowers: epistemic authority becomes attainable by all who can transcend everything that, by making them different, renders their judgments subjective, biased, and unreliable. Such transcendence has not, however, been available to all: since much of what is to be transcended is associated with the body, those whose lives have revolved around performing physical labor or tending to physical needs have not been encouraged to think of themselves as transcending that labor, nor have they had the leisure that Descartes correctly noted is needed to attain philosophical detachment and objectivity. Equally importantly, positions of privilege—such as maleness—have been invisible as positions, and those who have held them have not been marked as "different," while other positions have been marked as deviant, inferior, or otherwise epistemically tainting. The progress of liberalism has consisted largely in the expansion of the ranks of those who count as the same, whose differences have come to be at least officially regarded as merely superficial.

The fundamental challenge to liberalism and its associated epistemologies comes from a radical women's movement allied with radical movements in the Third World and among American people of color, along with others who are claiming authority from perspectives that are acknowledged to be irreducibly multiple, shifting, and, importantly, embodied. What is usually referred to as postmodernism, including, for example, the work of Richard Rorty, is largely a matter of sensitive seismic detection: some of those whose historical ground is shifting, because the bodies that have formed that ground are getting restless, have noticed that the world is not as stable as it once was. As feminists we need to keep in mind that our relationship to such academic postmodernism is that of the earthquake to the seismograph.

NAOMI SCHEMAN

University of Minnesota

THE POLITICS OF FRIENDSHIP*

my friends, there is no friend." In addressing you in this way, perhaps I have not yet said anything. Perhaps I have not even addressed myself to you. Not only because the two parts of this sentence, on the one side and on the other side of a comma or a pause, seem incompatible with each other, destined to annihilate themselves in their contradiction, but mainly because I have not yet said anything in my own name. I have contented myself with quoting. Spokesman for another, I have reported his words, which belong to a foreign or perhaps somewhat archaic language. I have, then, signed nothing, booked nothing to my own account.

"O my friends, there is no friend." This is not merely a citation which I am reading at present; it was already the quotation by another reader of the country I come from, Montaigne, "it is a saying which," he says, "Aristotle was used to repeating." In other words, I have quoted the quotation of a saying already quoted by Aristotle, a saying whose origin seems to lose itself in the anonymity of time immemorial. Nonetheless, it is not one of those proverbs without an assignable origin, and whose aphoristic mode rarely takes the form of an apostrophe.

By beginning in this way—quoting the quotation of a quotation—as I just said, I have perhaps not assumed the responsibility in my own name for any utterance. Perhaps I have not even yet addressed myself to you, really to you. But are things that simple? Am I completely irresponsible for what I have said when I am not responsible for what I have said? Am I not responsible for the fact that I have said (for the fact of having spoken) when I do not hold myself responsible for what I have said, for the content of what I have said and which I, in fact, have contented myself with reporting? Defined by what are commonly called conventions, a certain number of artificial signs attest to the following: even if I have not yet said anything determinate in my name when I uttered, in order to begin, without any further introduction, "O my friends, there is no friend," one has the right (but what is this right?) to suppose that I am nonetheless

^{*}To be presented in an APA symposium on Law and Society, December 30, 1988. Thomas A. McCarthy will comment; see this JOURNAL, this issue, 645–648. Given the constraints of time and place imposed on this publication, this is only the logical schema and minimal matrix of work in progress which will be published hereafter in an integral version.

^{1 &}quot;. . . il faut employer le mot qu'Aristote avoit tres-familier: O mes amis, il n'y nul amy." "De l'amitié," in *Essais*, Book I (Paris: Pléïade, 1959), p. 226.

speaking in my name. You hold me responsible, personally responsible, for the simple fact that I am speaking, and, for example, for the fact of quoting Montaigne in order to begin in place of and before saying anything else. And by holding me personally responsible, you are, in a rigorous sense, implying some knowledge of what 'person' and 'responsibility' mean.

What is happening at this very moment? This could give rise to a description of a "pragmatic" type. Such a description would confirm that, having been invited (but how and, exactly, by whom finally?) to speak to you when you are assembled to listen to me, then to discuss with me, in short, to respond to me, I have already responded to an invitation and, consequently, I am in the process of addressing myself to you who are beginning to respond to me. You are doing so in a way which is still virtual with respect to the content of the response, but you are already doing so actually with respect to that first response constituted by the attention given or at least promised to a discourse. (You should note in passing that, with this distinction between virtuality and act, I am already virtually installed in the dominant code, in the very constitution of one of the great canonical discourses of philosophy on friendship, the very one which Montaigne was quoting, Aristotle's. The distinction between dynamis and energeia is never far away, in the Nicomachean Ethics, when the issue is the distinction between the "good men who are friends in the rigorous sense of the term" and "the others who are so only accidentally and by analogy with the first" [VIII, ch. 4], or again when, after having defined the three "forms of government," Aristotle declares that "friendship appears [there] in the same proportion as justice," or, if man is a "political being" [IX, ch. 10], "political friendship" is only a kind of friendship, that which he calls "concord" [homonomia]. We should reconstitute these sequences.) You are already holding me responsible for what I say, for the simple fact that I am speaking, even if I am not yet assuming the responsibility for the sentences I am citing.

Supposing, concesso non dato, that one can translate these Greek words today by 'friendship', I still do not know if what exists between us is philia or homonomia, nor how one should distinguish here among us, among each one of us, who together would compose this as yet quite indeterminate "us." But perhaps you will concede the following, which is something like the first result of a practical demonstration, the one that has just taken place: before even having taken responsibility for any given affirmation, we are already caught up in a kind of asymmetrical and heteronomical curvature of the social space, more precisely, in the relation to the Other prior to any

organized socius, to any determined "government," to any "law."2 Please note: prior to any determined law, as either natural law or positive law, but not prior to any law in general. This heteronomical and asymmetrical curvature of a sort of originary sociality is a law, perhaps the very essence of the law. What is taking place at this moment, the disquieting experience we are having, is perhaps just the silent unfolding of that strange violence that has since forever insinuated itself into the origin of the most innocent experiences of friendship or justice. We have begun to respond. We are already caught, surprised [pris, surpris] in a certain responsibility, and the most ineluctable of responsibilities—as if it were possible to conceive of a responsibility without freedom. We are invested with an undeniable responsibility at the moment we begin to signify something (but where does that begin?). This responsibility assigns us our freedom without leaving it with us, if one could put it that way. And we see it coming from the Other. It is assigned to us by the Other, from the Other, before any hope of reappropriation permits us to assume this responsibility in the space of what could be called autonomy. This experience is even the one in which the Other appears as such, that is, appears without appearing.3 That which comes before autonomy must, then, also exceed it, that is, succeed it, survive it, and indefinitely surpass it. In general, when dealing with the law (nomos), one believes one can simply oppose autonomy and heteronomy. Perhaps one would have to deform this oppositional logic and prepare, from very far away, its "political" translation.

What can this excessive assignation of responsibility have to do with that which is called friendship? I say advisedly 'that which is called friendship', and I underscore this precaution. It resembles once again a quotation, as if I were forcing myself to remember unceasingly that, before knowing what friendship is and what we mean to say here and now with this word, we should first deal with a certain use of the word 'friendship'. We should mention these uses, as well as the interpretations and experiences (for experiences are also interpretations) that friendship has occasioned. For we should not forget that we are speaking first of all from within the tradition of a certain concept of friendship, within a given culture, let us say ours, in any case the one on the basis of which a certain "we" tries its luck here. Now, neither this tradition nor the concept of friendship within it is homogeneous. Our principal concern will even be to

² Here, of course, I am referring to the title of this panel.

³ Cf. my "Violence and Metaphysics," in Writing and Difference, Alan Bass, trans. (Chicago: University Press, 1978).

recognize the major marks of a tension within it, perhaps even ruptures, and, in any case, scansions.

Let us listen once again to Montaigne listening to Aristotle listening to the Other, but let us translate and interpret him as well: "O my friends, there is no friend." The painful and plaintive irony of the address also states the certitude of a strange affirmation. The phrase springs forth like a sort of apostrophe; in effect, someone is turning toward his friends, "O my friends . . . ," but the apostrophe carries within it a predicative proposition, it envelops an indicative declaration. Ascertaining a fact, it also utters a general truth: "there is no friend." The general truth of the fact would seem to contradict by an act the very possibility of the apostrophe, the possibility for it to be serious: there must indeed be friends in order for me to address myself to them in this way, if only so as to say to them "there is no friend." The contradiction would be as vivid and present as a simple logical absurdity; in the best of cases, it would be the playful exercise of a paradox, if the structures of the two utterances were symmetrical and if they belonged to a presently homogeneous ensemble. This is not necessarily the case. The apostrophe, whose form surpasses and comprises in itself the alleged determination of fact, resembles at one and the same time an act of recalling and an appeal [au rappel et à l'appel]. It resembles an appeal, because it makes a sign toward the future: be my friends, for I love or will love you (friendship, as Aristotle also said, consists rather in loving than in being loved [VIII 9, 25-30], a proposition on which we have not yet finished meditating), listen to me, be sensitive to my cry, understand and be compassionate; I am asking for sympathy and consensus, become the friends to whom I aspire. Accede to what is at the same time a desire, a request, a promise, and, one could also add, a prayer. And let us not forget what Aristotle said about prayer (eukhè): it is a discourse (logos), but it is a discourse that, somewhat in the manner of a performative, is neither true nor false [all'oúte alethès oúte pseudés]. 4 There are no friends, that we know, but I beg you, make it so that there will be friends from now on. What is more, how could I be your friend, and declare my friendship for you (and the latter consists more in loving than in being loved) if friendship did not remain something yet to happen, to be desired, to be promised? How could I give you my friendship where friendship would not be lacking, that is, if it already existed-more precisely, if the friend were not lacking? For the apostrophe does not say: "there is no friend-

⁴ Cf. my "Comment ne pas parler," in *Psyché: Inventions de l'autre* (Paris: Galilée, 1988), p. 572, n. 1.

ship," but rather "there is no friend." Perhaps this is because we have an idea of friendship and what it should be, in the ideality of its essence or telos, and thus in the name of friendship we must conclude, alas, that, if there is friendship, "there is no friend." (And this is just what Montaigne means to say in the context determined by the most thematic of his intentions, which dominates this passage up to a certain point: it is while thinking about "common friendships," "ordinary and customary" ones, that we are obliged to sigh with regret. These common friendships are not "the most perfect of their kind": that is why "there is no friend.") But, if there is no friend at present. then precisely let us make it so that there will be friendships from now on, friendships that are "the most perfect of their kind." Here is what I am calling you to, answer me, it is our responsibility. Friendship is never a given in the present; it belongs to the experience of waiting, of promise, or of commitment. Its discourse is that of prayer and at issue there is that which responsibility opens to the future.

But the apostrophe 'O my friends' turns also toward the past. It recalls, it makes a sign toward that which must be supposed so as to let oneself be understood, if only in the nonapophantic form of prayer. You have already shown me this minimal friendship, this preliminary consent without which you would not understand me, would not listen to my appeal, or be sensitive to what is hopeful in my cry. Without this absolute past, I could not, for my part, have addressed myself to you in this way. We would not be together in a sort of minimal community—but one which is also incommensurable with any other-speaking the same language or praying for translation within the horizon of the same language, even were it so as to manifest a disagreement, if a sort of friendship had not already been sealed before any other contract: a friendship prior to friendships, an ineffaceable, fundamental, and bottomless friendship, the one that draws its breath in the sharing of a language (past or to come) and in the being-together that any allocution supposes, including a declaration of war. Will one say, in a rather Aristotelian move, that this friendship has merely an accidental and analogical relation with friendship in the strict or proper sense, or with the friendship that is "perfect of its kind" (Montaigne)? The question thus becomes: "What is friendship in the proper sense?" "Is it ever present?" "What is the essence of friendship?" If we are not close to answering this question, it is not only because of the very great number of philosophical difficulties still in front of us and which we are going to try to approach. In a preliminary and principial manner, at the same time simple and abyssal, it is because the question 'what is?' (ti estin), the question of essence or truth, has unfolded itself, as the question

of philosophy, on the basis of a certain experience of *philein* and *philia*. The very possibility of the question, in the form of 'what is . . . ?', seems always to have supposed this friendship prior to friendships, this *anterior* affirmation of being-together in the allocution. Such an affirmation can no longer be simply integrated, above all it cannot be *presented* as a being-present (substance, subject, essence, or existence) within the space of an ontology, precisely because it opens this space.

Behind the logical game of contradiction or paradox, perhaps the 'O my friends, there is no friend' signifies first and last this surpassing of the present by the undeniable future anterior which would be the very movement and time of friendship. Undeniable future anterior, the absolute of an unpresentable past as well as future, which is to say of traces that one can only ever deny by summoning them into

⁵ There is not enough space to link this question to the elaboration of it which Martin Heidegger proposes, notably in Was ist das—Die Philosophie? (Pfüllingen: Neske, 1956). As we know, this elaboration also concerns the moment in which the philein of Heraclitus's philein to sophon, after having been determined as original accord (ein ursprünglicher Einklang, harmonia) would have become a tension toward searching, a jealous and tense inquisition (strebende Suchen) "determined by Eros." It is only with this eroticization of the questioning about being ("Was ist das Seiende, insofern es ist?") that thought (das Denken) would have become philosophy. "Heraclitus and Parmenides were not yet philosophers." The "step" toward philosophy would have been prepared by the Sophists and finally achieved by Socrates and Plato. Taking a careful reading of this interpretation as our guide, we might attempt to follow the very discreet thread of an incessant meditation on friendship in the path of Heidegger's thought. This meditation passes, in particular, by way of the unexpected and isolated allusion to the "friend's voice (Stimme des Freundes) that every Dasein carries within itself" (Being and Time, §34). Let us not forget that the existential analytics of Dasein, that "carries" (trägt) this voice in itself, is neither an anthropology, a sociology, nor an analytics of the subject, consciousness, psyche, or the self-it is neither a morals nor a politics. All these disciplines presuppose the analytics. This loads the allusion to the friend's voiceand thus to friendship itself—with a very particular ontological meaning, in a chapter on "Dasein und Rede, Die Sprache." This strange "voice," at once both internal and from elsewhere, has perhaps some relation to the "voice" of conscience (Gewissen) of which Heidegger also proposes an existential analytics (§57). Since the sex of this "friend" is not determined, I would also be tempted to graft onto this reading the questions I have elsewhere posed on the subject of the word Geschlecht and of sexual difference in Heidegger [cf. "Geschlecht" and "Geschlecht II" in Psyche, pp. 395-451, in Research in Phenomenology, XIII (1983), John P. Leavey, Jr., trans.; and in Deconstruction and Philosophy, John Sallis, ed. (Chicago: University Press 1987)]. These same questions should lead, by way of the Gespräch of the thinker with the poet, Gespräch that always supposes some sort of friendship, toward two types of texts: on the one hand, those which are addressed to Hölderlin ["Wo aber sind die Freunde?" in Andenken; cf. Heidegger's text which has the same title in Erlaüterungen zu Hölderlins Dichtung (Frankfurt: Klostermann, 1951)]; on the other hand, those which are addressed to Georg Trakl, to the figures of "the friend who follows the stranger," of the brother and sister, precisely around this motif of the Geschlecht ("Die Sprache im Gedicht," in Unterwegs zur Sprache [Pfüllingen: Neske, 1959]).

the light of phenomenal presence. A temporal torsion thus knots up the predicative proposition ('there is no friend') within the apostrophe ('O my friends'). The torsion of this asymmetry envelops the theoretical determination or the knowledge within the performativity of a prayer that it will never exhaust. This asymmetry leads us back to what I shall call the *question of the response*.

How should the question of the response be linked to the question of responsibility—and why should friendship be made into a privileged place for this reflection? A brief grammar of the response, or rather of "responding," will permit us a preliminary glimpse. I sketch such a grammar on the basis of my language, French, but I do not believe that, in this case, the concepts are thoroughly limited by language (even though idiomatic English here and there will require that we translate répondre by 'answer', the reader should not forget to hear the Latin root of 'responsibility'). Not that they are valid in general, beyond every language (syntax and lexicon), but I believe, in this context, they are translatable into the set of European languages which authorize us here to interrogate something such as our culture and our concept of responsibility. Which is to say that this grammar, however schematic, will be a bit more than a grammar.

One says "answer for," "answer to," "answer before" [répondre de, répondre à, répondre devant]. These three modalities are not juxtaposable; they envelop and imply each other. One answers for, for oneself or for something (for someone, for an action, for a thought, for a discourse), before, before another, a community of others, an institution, a tribunal, a law. And always one answers for, or before, by answering first to. This last modality thus appears more original, more fundamental, and hence unconditional.

1. One answers for oneself, for what one is, says, or does, and this beyond the simple present. The 'oneself' or 'myself' thus supposes the unity, in other words the memory, of the one responding. This is often called the unity of the subject, but one can conceive such a synthesis of memory without necessarily having recourse to the concept of subject. Since this unity is never secured in itself as an empirical synthesis, the recognition of this identity is entrusted to the instance of the name. "I" am held responsible for "myself," which is to say, for everything that can be imputed to that which bears my name. Imputability supposes freedom, to be sure, but it also supposes that that which bears my name remains the "same": not only from one moment to the next, from one state to the other of that which bears it, but also beyond even life or presence in general, for example, the presence to itself of that which bears it. The instance here of what is called the "proper name" is not necessarily limited to the phenome-

non of the legal name, the patronymic, or the social designation, although this phenomenon is, most frequently, its determining manifestation. We shall see that this question of the proper name is essential to the problematic of friendship. I find at least one indication in Montaigne's reflection. He says that his friendship for Etienne de la Boétie preceded their meeting. More precisely, this meeting or acquaintance [accointance] took place long "before I had seen him, and gave me the first knowledge of his name, thus leading this friendship on its way."

There is, beyond my whole discourse, and what I can specifically say about it, some unknown, inexplicable and fatal force, the go-between of our union. We sought each other before we had seen one another, and through the reports we heard about each other, which caused a greater striving in our feelings than that occasioned by the sense of the reports, I believe through some ordinance of heaven: we embraced each other through our names ("De l'amitié," p. 225).

- 2. One answers first to the Other: to the question, the request, the prayer, the apostrophe, the appeal, the greeting, or the sign of the Other. This dimension of responding, as responding to, is more original than the others, as we have noted, for two reasons. On the one hand, one does not answer for oneself and in one's own name, one is not responsible except before [devant] the question, request, challenge, "instance," or "insistance" of the Other. On the other hand, the proper name that structures the "answering for oneself" is in itself for the Other, whether because the Other has chosen it (for example, the name I am given at birth, which I never chose and which introduces me into the space of the law-and the law is the theme of this session), or whether because, in any case, it implies the Other in the very act of naming, its origin, its finality, its use. Responding always supposes the Other in the relation to oneself; it preserves the sense of this asymmetrical "anteriority" even within the seemingly most inward and solitary autonomy of reserve [quant \hat{a} soi], of one's heart of hearts, and of the moral conscience jealous of its independence—another word for freedom. This asymmetrical anteriority also marks temporalization as a structure of responsibility.
- 3. Answering before: this expression seems at first to modalize 'answering to'. One answers before the Other because first one answers to the Other. But this modalization is more than or other than a specification by example. A decisive round is being played out here, and we should record all its effects. In the idiom, the expression 'before' generally indicates the passage to an institutional instance of

alterity. It is no longer singular, but is universal in its principle. One answers to the Other who can always be singular, and who must remain so in a certain way, but one answers before the law,6 a tribunal, a jury, some agency (instance) authorized to represent the Other legitimately, in the form of a moral, legal, or political community. Here we have two forms or two dimensions of the respect implied by any responsibility (I note in passing that these two words. 'respect' and 'responsibility', which are linked and constantly incite each other, appear to refer in the first case to distance, to space, and to the look [regard], and, in the second case, to time, to the voice, and to listening. Their co-implication can be sensed at the heart of friendship, one of whose enigmas comes from this distance or this respectful separation which distinguishes it, as a feeling, from love. This co-implication calls for a rigorous rereading of the Kantian analysis of respect in friendship. There is no friendship without "respect of the Other," but this respect, although inseparable from a "morally good will," should not be simply confused with purely moral respect, the respect owed only to its "cause," the moral law, of which the person is but an example.)

Of these two dimensions of the relation to the Other, the one maintains the absolute singularity of the Other and of "my" relation to the Other, as well as the relation of the Other to the Other which I am for him. But the relation to the Other also passes through the

⁶ Cf. my "Devant la loi," in La Faculté de juger (Paris: Minuit, 1985) [in Kafka and the Contemporary Critical Performance, A. Ronell, trans., A. Udoff, ed. (Bloomington: Indiana UP, 1987)] and Parages (Paris: Galilée, 1986).

7 Cf. Foundations of the Metaphysics of Morals. Here I can only indicate the program of a reading. The principal text that Kant devotes to friendship is immensely complex. It constitutes precisely the "Conclusion of the Elementary Doctrine" (of ethics) in the "Doctrine of Virtue" (The Metaphysics of Morals, §46/7). Kant also quotes there, in a slightly different form, Aristotle's saying ("My dear friends, there exist no friends!"). In its perfection, that is, as an unrealizable but practically necessary Idea, friendship supposes both love and respect. It must be equal and reciprocal. To seek it is a duty because, although friendship does not produce happiness, the two feelings that compose it envelop the "dignity of being happy." But one of the difficulties, in the very idea of friendship, comes from the contradictory character and thus the unstable equilibrium of the two feelings that are opposed in the mode of "attraction" that tends toward fusion (love) and "repulsion" that holds at a distance (respect). A reflection on the Kantian ethics and politics of friendship should be organized around the concept of "secret." It seems to me to dominate §47 and to mark, in a problematic way, the ideal of friendship as communication (Mitteilung) and egalitarian sharing. Such a reflection would consider first the definition of the "friend of mankind." Every friend must first of all be the "friend of mankind." The latter is not only the philanthropist. The friend of mankind supposes equality among men, the idea of being obligated by this very equality. Is it just by chance that the familial schema imposes itself once again here, and in these terms (father/brothers)? "Here one represents to oneself all men as brothers submissive to a universal father who wants the happiness of all."

universality of the law. This discourse about universality which can find its determination in the regions of morality, law, or politics, always appeals to a third party, beyond the face-to-face of singularities. The third party is always witness for a law that comes along to interrupt the vertigo of singularity. Do we have here two models of friendship, of which the one would find its motto in Montaigne's response ("If one presses me to say why I loved him, I feel that can only be expressed by responding: Because it was he; because it was me",—ob. cit.); and the other in Zarathustra's sentences when it is a question of interrupting the jealous narcissism of the dual relation, which always remains enclosed between "me" and "me," "I" and "me." of preventing it from sinking into the abyss ("I and me are always in too jealous [too zealous, zu eifrig] a dialogue: how could one endure this situation if there were no friend? For the hermit, the friend is always the third party: the third party is the cork that prevents the dialogue of the two from sinking into the abyss."8 But is this an alternative? Are there really two different, even antagonistic or incompatible relations? Do not these two relations imply each other at the moment they seem to exclude each other? Does not my relation to the singularity of the Other as Other pass through the law? Does not the law command me to recognize the transcendent alterity of the Other who can only ever be heterogeneous and singular, hence resistant to the very generality of the law? But this co-implication, far from dissolving the antagonism and breaking through the aporia, aggravates them instead—at the very heart of friendship.

Sharing [le partage] (singularity/universality) has always divided the experience, the concept, and the interpretation of friendship. It has determined other oppositions there (secret, private, invisible, unreadable, apolitical, or even without a concept versus manifest, public, exposed to witnesses, political, homogeneous with the concept). Between the two terms of the opposition, there is the familial schema (I am using the word 'schema' in the Kantian sense: between intuitive singularity and the generality of the concept). On the one hand, friendship seems to be essentially foreign or unamenable to the res publica and thus could not found a politics. But, on the other hand, as one knows, from Plato to Montaigne, from Aristotle to Kant, from Cicero to Hegel, the great philosophical and canonical discourses on friendship (but my question goes precisely to the philo-

^{8 &}quot;Ich und Mich sind immer zu eifrig im Gespräche: wie wäre es auzuhalten, wenn es nicht einen Freund gäbe?

[&]quot;Immer ist für den Einsiedler der Freund der Dritte: der Dritte ist der Kork, der verhindert, dass das Gespräch der Zweie in die Tiefe sinkt." "Dem Freunde" in Also sprach Zarathustra.

sophical canon in this domain) will have linked friendship explicitly to virtue and to justice, to moral reason and to political reason. These discourses will have even set the moral and political conditions for an authentic friendship—and vice-versa. Obviously, these discourses differ among themselves and would call for long and careful analyses. Such analyses should take care, in particular, not to identify too quickly morality with politics, in the name of the law: it is sometimes in the name of morality that friendship has been removed from the divisions and criteria of politics.

These oppositions seem to dominate the interpretation and the experience of friendship in our culture: a domination which is unstable and under internal stress, but which is therefore all the more imperious. What relation does this domination maintain with the double exclusion that can be seen at work in all the great ethicopolitico-philosophical discourses on friendship, namely, on the one hand, the exclusion of friendship between women, and, on the other hand, the exclusion of friendship between a man and a woman? This double exclusion of the feminine in the philosophical paradigm of friendship would thus confer on it the essential and essentially sublime figure of virile homosexuality. Within the familial schema, whose necessity I mentioned earlier, this exclusion privileges the figure of the brother, the name of the brother or the name of brother, more than that of the father—whence the necessity of connecting the political model, especially that of democracy and of the Decalogue, with the rereading of Freud's hypothesis about the alliance of brothers.9 Again Montaigne on his friendship with La Boétie: "In truth, the name of brother is a beautiful and delectable one, and for this reason we made it, he and I, our alliance."

These exclusions of the feminine would have some relation to the movement that has always "politicized" the model of friendship at the very moment one tries to remove this model from an integral politicization. The tension here is within politics itself. It would be necessary to analyze all discourses that reserve politics and public space for man, domestic and private space for woman. For Hegel, this is also the opposition between day and night, and hence a certain number of other oppositions as well. What is Nietzsche's place in this "history"? Does he profoundly corroborate an old tradition ("That is why woman is not yet capable of friendship: she only knows

⁹ I underscored the difficulties and the paradoxes of the Freudian hypothesis in

¹⁰ On all these problems and once again on the ethico-political question of the woman, the sister, and the brother in Hegel, see my *Glas* (Paris: Galilée, 1974) [John P. Leavey, Jr. and Richard Rand, trans. (Lincoln: Nebraska UP, 1987)].

love")?¹¹ Or is he trying to think a friendship that goes beyond Judeo-Christian and philosophical history ("I am not teaching you the neighbor [das Nächsten] but rather the friend. Let the friend be for you the feast of the earth and the foreshadow of the superhuman. I am teaching you the friend and his overflowing heart")?¹²

If the great canonical meditations on friendship (for example, Cicero's De amicitia, Montaigne's De l'amitié, Maurice Blanchot's L'amitié) are linked to the experience of mourning, to the moment of loss—that of the friend or of friendship—if, through the irreplaceable element of the name, they always advance into the testamentary shadow in order to entrust and refuse the death of the unique one to a universalizable discourse [". . . my friends, there is no friend" (Aristotle-Montaigne), "But what have become of the friends?" (Villon), "Wo aber sind die Freunde?" (Hölderlin)], if they thereby found and destabilize at the same time, if, because they menace them, they restore a great number of oppositions (singular/ universal, private/public, familial/political, secret/phenomenal, etc.) and I would be tempted to say all oppositions, the relative invariance of this model fractures itself and opens itself onto its own abyss. By returning to all motifs I have just sketched (the morals and the politics of friendship, death, the name, fraternity, etc.), by reconsidering all the oppositions I have just situated, I would have liked to try to recognize two major ruptures in what one could, as a matter of convenience, call the history of friendship. (But a certain friendship could make the most traditional concept of historicity quake.) The

^{11 &}quot;Deshalb ist das Weib noch nicht der Freundschaft fähig: es kennt nur die Liebe." One must underscore here the 'not yet', because it also extends to man (Mann), but first of all and once again to the "brother" of Zarathustra as to the future of a question, an appeal or a promise, a cry or a prayer. It does so in the performative mode of the apostrophe. There is as yet no friendship, no one has yet begun to think friendship. Nevertheless, in the experience of a sort of bereaved anticipation, we can already name the friendship that we have not yet encountered. We can already think that we do not yet have access to it. May we be able to do it! That is the exclamation point, the singular clamor of this "claim." Here is the "Oh my friends, there is no friend" of Zarathustra: "Woman is not yet capable of friendship. But tell me, men, who among you is capable of friendship? . . . There is camaraderie: may there be friendship!" (Aber sagt mir, ihr Männer, wer von euch ist denn fähig der Freundschaft? . . . Es gibt Kameradschaft: möge es Freundschaft geben!). But since woman has not yet acceded to friendship because she remains—and that is love—either "slave" or "tyrant," the friendship to come continues to mean for Zarathustra: liberty, equality, fraternity. In short, the motto of a republic.

[&]quot;Nicht des Nächsten lehre ich euch, sondern den Freund. Der Freund sei euch das Fest der Erde und ein Vorgefühl des Uebermenschen.

[&]quot;Ich lehre euch den Freund und sein übervolles Herz" ("Von des Nächstenliebe"). With the love of the distant one, Zarathustra advises love of the future—and beyond humanity, love of things and fantoms (die Liebe zu Sachen und Gespenstern).

Greco-Roman model appears to be marked by the value of reciprocity, by homological, immanentist, finitist, and politicist concord. Montaigne (whom we are reading here as the example of a paradigm) doubtless inherits the majority of these traits. But he breaks the reciprocity therein and discreetly introduces, so it seems to me, heterology, asymmetry, and infinity ("he surpassed me by an infinite distance;" "I would have certainly entrusted myself more willingly to him than to me"; "For even the discourses which Antiquity has left us on this subject seem to me to be slack in seizing the feeling which I have about it").

Shall one say that this fracture is Judeo-Christian? Shall one say that it depoliticizes the Greek model or that it displaces the nature of the political? Can the same type of question be put regarding Nietzsche and Blanchot (other examples where friendship should defy both historicity and exemplarity)? In a different way, to be sure, both call the friend by a name that is no longer that of the neighbor [prochain], perhaps no longer that of a man. 13

JACQUES DERRIDA

École des Hautes Études [Translated by Gabriel Motzkin]

13 The "Who?" of friendship moves off into the distance beyond all these determinations. In its "infinite imminence," it even exceeds the interest of knowledge, science, truth, proximity, even life and even the memory of life. It is not yet an

identifiable "I," either public or private.

"We ought to renounce trying to know those to whom we are linked by something essential; I mean, we ought to welcome them in the relation with the unknown where they welcome us, as well, in our distance. Friendship, that relation without dependence, without episode and yet into which enters all the simplicity of life, passes by way of the recognition of the common strangeness that does not allow us to speak of our friends, but only to speak to them; it does not allow us to make them the theme of conversations (or of articles), but is the movement of the understanding in which, speaking to us, they keep, even in moments of the greatest familiarity, their infinite distance, that fundamental separation on the basis of which that which separates become relation. Here, discretion is not in the simple refusal to refer to what one has learned in confidence (how disgraceful that would be, even just to think about it), but it is the interval, the pure interval that, from me to this other who is a friend, measures everything there is between us, the interruption of being that never authorizes me to dispose of him, nor of my knowledge of him (even were it to praise him) and that, far from preventing any communication, relates us to one another in the difference and sometimes the silence of speech [parole]." At the death of the friend, the "measureless movement of dying," "the event" of death reveals and erases at the same time this "truth" of friendship: "no, not the deepening of the separation, but its effacement; not the enlarging of the caesura, but its leveling and the dissipation of that void between us where, long ago, there developed the frankness of a relation without history. The result is that, at present, what was close to us not only has ceased to approach, but has lost even the truth of the extremely distant. . . . We can, in a word, remember. But thought knows that one does not remember: without memory, without thought, it already struggles in the

Digitized by Arya Samaj Foundation Chennal and eGangotri ON THE MARGINS OF POLITICS*

THE general question raised by Jacques Derrida's subtle variations on Levinasian and Heideggerian themes is how best to be postmetaphysical in thinking about ethics, law, and politics. Is it by pursuing a deconstructionist strategy that remains at the level of metaphysics in order to disrupt and displace it? I think not, and want to indicate why by (all too briefly) suggesting another reading of the phenomena he alludes to, one that draws on just those domains of social research and practical philosophy which Heidegger devalued in relation to ontology (cf. fn. 5). A more specific question is whether "friendship" is the best place to start thinking about ethics, law, and politics. It might seem so if one's overriding concern is to escape the conceptual compulsion to identify (being with thought, the individual with the universal, etc.) which drives traditional metaphysics. I want to argue that there are other, less metaphysically motivated, ways of thinking about social relations so as not to exclude or forcefully assimilate what is different.

We become individuals in and through being socialized into shared forms of life. In this sense, the "Other" (here: culture, society) does indeed "come before autonomy." Because we become who we are by growing into a network of social relations, we are always "already caught up in . . . a curvature of social space." This is, however, "prior to any organized socius" only in the sense that the lifeworld is prior to formal law and formally organized spheres of life. Everyday life is itself highly organized and structured. And, as Harold Garfinkel and others have made clear, it is our shared understanding of social structure which informs the normative expectations we bring to social situations. (To pick up on Derrida's example: it is our knowledge of what it means to be invited to address a scholarly organization, to assemble for that purpose, to present a formal response to a

realm of the invisible where everything falls back into indifference. That is its profound grief. It must accompany friendship into oblivion." [Maurice Blanchot, L'Amitié (Paris: Gallimard, 1971), pp. 326–330.] The epigraph to the book are these words of Georges Bataille: "friends even to that state of profound friendship where a man abandoned, abandoned by all his friends, meets up in life with the one who will accompany him beyond life, himself without life, capable of free friendship, detached from any ties."

* Abstract of a paper to be presented in an APA symposium on Law and Society, December 30, 1988, commenting on a paper by Jacques Derrida, this JOURNAL, this issue, 632-644. All phrases appearing in quotation marks are citations from that

0022-362X/88/8511/0645\$00.50

© 1988 The Journal of Philosophy, Inc.

paper, etc., which informs what we are doing here). We orient our behavior to these shared schemes of interpretation and expectation, mutually attributing understanding of them to one another and holding one other responsible in terms of them. In short, at the level of our everyday interactions we normally believe ourselves to be, and take others as being, knowledgeable subjects confronted by real choices, for which we and they will be held accountable.

But this means that the network of reciprocally connected and temporally generalized expectations that George Herbert Mead called the generalized Other has a mechanism for individuation built right into it. It is the individual actor who must "answer to the Other," meet or fail to meet the others' expectations, and in more or less standard or novel ways, who can answer "yes" or "no" to the claims made by others to and upon her. Thus, it is not only in friendship, but in social interaction generally, that the "singularity of the Other" is intimately interconnected with the "generality of the law" (here: normative expectations). And although the individual is related "asymmetrically" to the generalized Other, which is always "anterior," the socially generalized patterns of behavior the latter comprises are themselves typically structured as relations of reciprocity with individual Others: the actor is entitled to expect certain kinds of behavior from others in certain situations, and is obligated in turn to meet their legitimate expectations. Moreover, we are typically expected to justify behavior that deviates from what is legitimately expected. From this perspective, what is built into "the sharing of a language" and "being together" is not so much a "minimal friendship" as a minimal ethics of reciprocity and accountability. Moreover, inasmuch as individuation and socialization are two aspects of the same process, personal identity is interwoven with a fabric of relations of mutual recognition. This brings with it a reciprocal vulnerability that calls for guarantees of mutual consideration to preserve both the integrity of individuals and the web of interpersonal relations in which they form and stabilize their identities. Both of these concerns—with the dignity and inviolability of the person, and with the good of the community as a whole—have been at the heart of traditional moralities. It is our intuitive grasp of the minimal ethics built into the structure of social interaction and of the complementary role played by morality which philosophical ethics seeks reflectively to articulate, refine, and elaborate.

In the Kantian tradition, equal respect has been tied to autonomy —the freedom of moral subjects to act upon laws they themselves accept as binding on the basis of their own insight—and concern for the common good to the impartiality of laws that could be accepted

by everyone on that basis. Contemporary neo-Kantians typically retain this dual thrust, but shift the frame of reference from Kant's solitary reflecting moral consciousness to the community of moral subjects in dialogue. In Jürgen Habermas's discourse ethics, for example, whether a norm is justified cannot be determined monologically but only by discursively testing its claim to be justified through the informed, uncoerced, reasoned agreement of all those subject to it. In this representation of the moral point of view, equal respect for individuals is reflected in the requirement of rationally motivated agreement, such that each has the right to respond with a "yes" or "no" to the *reasons* offered by way of justification; concern for the common good is reflected in the requirement of general and reciprocal perspective taking, such that, in seeking mutual agreement, each attempts to get beyond the egocentric viewpoint by taking the situations of others into account and giving them equal weight to his own.

This is of course a highly idealized account of moral and legal relations, and so it is meant to be. In the Kantian tradition, regulative ideas are marked by the same sort of "promise," "commitment," and "responsibility to the future" that Derrida sees in the notion of "friendships that are 'the most perfect of their kind'." Although never fully actual, regulative ideas are nonetheless actually effective in structuring our practices—in guiding our efforts to fashion just laws and institutions, in shaping our perceptions and criticisms of injustice, and so forth. It is important to keep in mind here that Kant introduced his regulative ideas precisely by way of contrast to the constitutive ideas of the Platonic tradition. Nothing in experience can correspond to regulative ideas; they are not representable in and of themselves, but only in relation to the practices they regulate. Thus, Derrida's subversive reaction to the overblown claims of the Western theoretical tradition is less well-suited to thinking about the always imperfect practice of justice.

The message intended by this brief sketch of an alternative to "the politics of friendship" is that "the curvature of social space" is not fundamentally "asymmetrical and heteronomical"; "originary sociality" is marked as well by relations of symmetry, reciprocity, and mutual recognition. Mastering such relations belongs to the repertoire of competent social actors—e.g., to the competence to play social roles of any sort; and this entails an intuitive mastery of the rudiments of the moral point of view. It is true that general norms justified from the standpoint of impartial judgment will require abstracting from the concrete circumstances of specific cases, and in this sense will necessarily be unheeding of difference and in violation of singularity. But it should also be noted that the point of view of

fairness is at the same time respectful of the Other and tolerant of difference; it functions precisely to place limits on egocentric violations of the Other by forcing a de-centering of perspective. Furthermore, the process of applying general norms fairly itself requires at least a partial reversal of the abstractions required to justify them—for example, by requiring attention to the specifics of individual cases to determine whether they are really alike in relevant respects.

Treating friendship as "a privileged place for reflecting" on law, morality, and politics does provide an instructive antidote to the leveling, difference-denying tendencies of much moral and legal theory, of the Kantian sort as well. It helps make us attentive to what gets left out of generalizing schemes, of what is ignored, excluded, or assimilated by them. And it can serve to remind us that universalizing theories of justice do not exhaust the domain of the ethical. But Derrida's basic motivation for taking it as paradigmatic lies, I think, elsewhere—namely, in the link he sees between "reciprocity" and the "homological, immanentist, and finitist" tendencies of Western logocentrism, whose dominance the practice of deconstruction is designed precisely to disrupt. In this context, friendship can serve as an allegory for what gets left out of totalizing systems of representation, for "heterogeneity, asymmetry, and infinity." The costs of approaching law by this route are very high, however; the "heteronomical" can appear to be its "very essence," and any attempt to name the law to be a violation of an ancient proscription. I have tried to suggest that a better way of being postmetaphysical in ethics, law, and politics is to stop doing metaphysics, even of a negative sort, when thinking about them. If we do that, talk of reciprocity, symmetry, mutual recognition, and the like will no longer seem to be a hybristic attempt to say the unsayable, but rather a modest effort to explicate the very ideas of respect, tolerance, and solidarity which antiuniversalist invocations of Otherness and difference generally take for granted.

Northwestern University

THOMAS A. McCARTHY

INDIVIDUALISM AND SELF-KNOWLEDGE*

HE problem I want to discuss derives from the juxtaposition of a restricted Cartesian conception of knowledge of one's own thoughts and a nonindividualistic conception of the individuation of thoughts. Both conceptions are complex and controversial. But I shall not explain them in detail, much less defend them. I shall explicate them just enough to make the shape of the problem vivid. Then I shall say something about solving the problem.

Descartes held that we know some of our propositional mental events in a direct, authoritative, and not merely empirical manner. I believe that this view is correct. Of course, much of our self-knowledge is similar to the knowledge of others' mental events. It depends on observation of our own behavior and reliance on others' perceptions of us. And there is much that we do not know, or even misconstrue, about our own minds. Descartes tended to underrate these points. He tended to overrate the power of authoritative self-knowledge and its potential for yielding metaphysical conclusions. Characterizing the phenomenon that interested Descartes is a substantial task. I shall not take on this task here. I think, however, that Descartes was right to be impressed with the directness and certainty of some of our self-knowledge. This is the point I shall rely on.

Descartes's paradigm for this sort of knowledge was the cogito. The paradigm includes not only this famous thought, but fuller versions of it—not merely 'I am now thinking', but 'I think (with this very thought) that writing requires concentration' and 'I judge (or doubt) that water is more common than mercury'. This paradigm goes further toward illuminating knowledge of our propositional attitudes than has generally been thought. But I note it here only to emphasize that Descartes's views about the specialness of some self–knowledge are not merely abstract philosophical doctrine. It is certainly plausible that these sorts of judgments or thoughts constitute knowledge, that they are not products of ordinary empirical investigation, and that they are peculiarly direct and authoritative. Indeed, these sorts of judgments are self-verifying in an obvious way: making these judgments itself makes them true. For mnemonic purposes, I shall call such judgments basic self-knowledge.

0022-362X/88/8511/0649\$01.50

© 1988 The Journal of Philosophy, Inc.

^{*}To be presented in an APA symposium on Individuation and Self-Knowledge, December 30, 1988. Donald Davidson will comment; see this JOURNAL, this issue, 664/5. Substantially this paper was the Nelson Lecture, University of Michigan, February 1986. I benefited from the occasion.

Let us turn from knowledge of one's thoughts to individuation of one's thoughts. My view on this matter is that many thoughts are individuated nonindividualistically: individuating many of a person or animal's mental kinds—certainly including thoughts about physical objects and properties—is necessarily dependent on relations that the person bears to the physical, or in some cases social, environment. This view is founded on a series of thought experiments, which I shall assume are familiar. Their common strategy is to hold constant the history of the person's bodily motion, surface stimulations, and internal chemistry. Then, by varying the environment with which the person interacts while still holding constant the molecular effects on the person's body, one can show that some of the person's thoughts vary. The details of the thought experiments make it clear that the variation of thoughts is indicative of underlying principles for individuating mental kinds. The upshot is that which thoughts one has-indeed, which thoughts one can have-is dependent on relations one bears to one's environment.

Our problem is that of understanding how we can know some of our mental events in a direct, nonempirical manner, when those events depend for their identities on our relations to the environment. A person need not investigate the environment to know what his thoughts are. A person does have to investigate the environment to know what the environment is like. Does this not indicate that the mental events are what they are independently of the environment?

By laying aside certain contrary elements in Descartes's views, one can reconstruct a tempting inference to an affirmative answer from his conception of self-knowledge.

In reflecting on the demon thought experiment, one might think that, since we can know our thoughts authoritatively, while doubting whether there is any physical world at all, the natures of our thoughts—our thought kinds—must be independent of any relation to a physical world. A parallel inference is presupposed in Descartes's discussion of the real distinction between mind and body. In *Meditations* VI, he argues that the mind can exist independently of any physical entity. He does so by claiming that he has a "clear and distinct idea" of himself as only a thinking and unextended thing,

¹ Cf. my "Individualism and the Mental," Midwest Studies in Philosophy, 1V (1979): 73–121; "Other Bodies," in Thought and Object, Andrew Woodfield, ed. (New York: Oxford, 1982); "Individualism and Psychology," The Philosophical Review, XCV, 1 (1986): 3–45; "Cartesian Error and the Objectivity of Perception," in Subject, Thought, and Context, Philip Pettit and John McDowell, eds. (New York: Oxford, 1986); "Intellectual Norms and Foundations of Mind," this JOURNAL, LXXXIII, 12 (December 1986): 697–720.

and a "clear and distinct idea" of body as only an extended and unthinking thing. He claims that it follows that the mind that makes him what he is can exist independently of any physical body. The argument also occurs in *Principles* 1, LX:

. . . because each one of us is conscious [through clear and distinct ideas] that he thinks, and that in thinking he can shut off from himself all other substance, either thinking or extended, we may conclude that each of us . . . is really distinct from every other thinking substance and from every corporeal substance.²

Descartes also believed that he had "clear and distinct ideas" of his thoughts. One might argue by analogy that, since one can "shut off" these thoughts from all corporeal substance, they are independent for their natures from physical bodies in the environment, and presumably from other thinkers. This line of argument implies that knowledge of one's own thoughts guarantees the truth of individualism.³

The root mistake here has been familiar since Arnauld's reply. It is that there is no reason to think that Descartes's intuitions or self-knowledge give him sufficient clarity about the nature of mental events to justify him in claiming that their natures are independent of relations to physical objects. Usually, this point has been made against Descartes's claim to have shown that mental events are independent of a person's body. But it applies equally to the view that mental kinds are independent of the physical environment. One can know what one's mental events are and yet not know relevant general facts about the conditions for individuating those events. It is simply not true that the cogito gives us knowledge of the individuation conditions of our thoughts which enables us to "shut off" their individuation conditions from the physical environment. Our thought experiments, which have directly to do with conditions for individuation, refute the independence claim.⁴

It is one thing to point out gaps in inferences from self-knowledge to individualism. It is another to rid oneself of the feeling that there is a puzzle here. Why is our having nonempirical knowledge of our

² The Philosophical Works of Descartes, vol. 1, Haldane and Ross trans. (New York: Dover, 1955), pp. 243/4.

³ Cf. *ibid.*, p. 190.
⁴ I have discussed this and other features of the inference in "Cartesian Error and the Objectivity of Perception." See also my "Perceptual Individualism and Authoritative Self-Knowledge," in *Contents of Thought*, Robert Grimm and Daniel Merrill, eds. (Tucson: Arizona UP, 1988). I now think that Descartes's views have more anti-individualistic elements than I realized in writing those articles. I hope to discuss these matters elsewhere.

thoughts not impugned by the fact that such thoughts are individuated through relations to an environment that we know only empirically?

Let us assume that our thoughts about the environment are what they are because of the nature of entities to which those thoughts are causally linked. According to our thought experiments, a person with the same individualistic physical history could have different thoughts if the environment were appropriately different. One senses that such a person could not, by introspection, tell the difference between the actual situation (having one set of thoughts) and the counterfactual situation (having another).

This intuition must be articulated carefully. What do we mean by 'introspection'? In each situation, the person knows what his thoughts are; and in each situation the thoughts are different. If 'introspection' were explicated in terms of self-knowledge, there would be an introspectible difference.

Certainly, if one were stealthily shifted back and forth between actual situations that modeled the counterfactual situations, one would not notice some feature in the world or in one's consciousness which would tell one whether one was in the "home" or the "foreign" situation. But this remark does not capture the idea that the two lives would feel the same. The thoughts would not switch as one is switched from one actual situation to another twin actual situation. The thoughts would switch only if one remained long enough in the other situation to establish environmental relations necessary for new thoughts. So quick switching would not be a case in which thoughts switched but the introspection remained the same.

But slow switching could be such a case. Suppose that one underwent a series of switches between actual earth and actual twin earth so that one remained in each situation long enough to acquire concepts and perceptions appropriate to that situation. Suppose occasions where one is definitely thinking one thought, and other occasions where one is definitely thinking its twin. Suppose also that the switches are carried out so that one is not aware that a switch is occurring. The continuity of one's life is not obviously disrupted. So, for example, one goes to sleep one night at home and wakes up in

⁵ Of course, there can arise difficult questions about whether one is still employing thoughts from the departed situation or taking over the thoughts appropriate to the new situation. I think that general principles govern such transitions, but such principles need not sharply settle all borderline cases. Insofar as one finds problems associated with actual switches distracting, one could carry out the objection I am articulating in terms of counterfactual situations.

twin home in twin bed—and so on. (Your standard California fantasy.) Now suppose that, after decades of such switches, one is told about them and asked to identify when the switches take place. The idea is that one could not, by making comparisons, pick out the twin periods from the "home" periods.

I grant these ideas. The person would have no signs of the differences in his thoughts, no difference in the way things "feel." The root idea is that at least some aspects of one's mental life are fixed by the chemical composition of one's body. One might call these aspects pure phenomenological feels. If one were uncomfortable with this notion, one could explicate or replace it in terms of an abstraction from the person's inability to discriminate between different mental events under the stated switching situations.

The upshot of all this is that the person would have different thoughts under the switches, but the person would not be able to compare the situations and note when and where the differences occurred. This point easily, though I think mistakenly, suggests the further point that such a person could not know what thoughts he had unless he undertook an empirical investigation of the environment which would bring out the environmental differences. But this is absurd. It is absurd to think that, to know which thoughts we think, we must investigate the empirical environment in such a way as to distinguish our actual environment from various twin environments.

In basic self-knowledge, a person does individuate his thoughts in the sense that he knows the thought tokens as the thought tokens, and types, that they are. We know which thoughts we think. When I currently and consciously think that water is a liquid, I typically know that I think that water is a liquid. So much is clear.

How can one individuate one's thoughts when one has not, by empirical methods, discriminated the empirical conditions that determine those thoughts from empirical conditions that would determine other thoughts?

It is uncontroversial that the conditions for thinking a certain thought must be presupposed in the thinking. Among the conditions that determine the contents of first-order empirical thoughts are some that can be known only by empirical means. To think of something as water, for example, one must be in some causal relation to water—or at least in some causal relation to other particular substances that enable one to theorize accurately about water. In the normal cases, one sees and touches water. Such relations illustrate the sort of conditions that make possible thinking of something as water. To know that such conditions obtain, one must rely on empiri-

cal methods. To know that water exists, or that what one is touching is water, one cannot circumvent empirical procedures. But to *think* that water is a liquid, one need not *know* the complex conditions that must obtain if one is to think that thought. Such conditions need only be presupposed.

Now let us turn to knowledge of one's thoughts. Knowing what one is thinking when one has thoughts about physical entities presupposes some of the same conditions that determine the contents of the empirical thoughts one knows one is thinking. This is a result of the second-order character of the thoughts. A knowledgeable judgment that one is thinking that water is a liquid must be grounded in an ability to think that water is a liquid.

When one knows that one is thinking that p, one is not taking one's thought (or thinking) that p merely as an object. One is thinking that p in the very event of thinking knowledgeably that one is thinking it. It is thought and thought about in the same mental act. So any conditions that are necessary to thinking that p will be equally necessary to the relevant knowledge that one is thinking that p. Here again, to think the thought, one need not know the enabling conditions. It is enough that they actually be satisfied.

Both empirical thoughts and thinking that one is thinking such thoughts presuppose conditions that determine their contents. In both cases, some of these conditions can be known to be satisfied only by empirical means. Why do these points not entail that one cannot know that one is thinking that such and such unless one makes an empirical investigation that shows that the conditions for thinking such and such are satisfied? The answer is complex, but it can be seen as a series of variations on the point that one must start somewhere.

It is helpful in understanding self-knowledge to consider parallel issues regarding perceptual knowledge. It is a fundamental mistake to think that perceptual knowledge of physical entities requires, as a precondition, knowledge of the conditions that make such knowledge possible. Our epistemic right to our perceptual judgments does not rest on some prior justified belief that certain enabling conditions are satisfied. In saying that a person knows, by looking, that there is food there, we are not required to assume that the person knows the causal conditions that make his perception possible. We certainly do not, in general, require that the person has first checked that the light coming from the food is not bent through mirrors, or that there is no counterfeit food in the vicinity. We also do not require that the person be able to recognize the difference between

food and every imaginable counterfeit that could have been substituted.

In fact, it is part of our common conception of the objectivity of perception that there is no general guarantee that the perceiver's beliefs, dispositions, and perceptions could in every context suffice to discriminate the perceived object from every possible counterfeit. The possibility of unforeseeable misperceptions and illusions is fundamental to objectivity. So the very nature of objective perception insures that the perceiver need not have a perfect, prior mastery over the conditions for his perceptual success.

This point is obvious as applied to common practice. But it is the business of philosophy and the pleasure of skepticism to question common practice. My discussion of knowledge and individualism has proceeded on the unargued assumption that skepticism is mistaken. Granted this assumption, the point that perceptual knowledge does not require knowledge of its enabling conditions is obvious.

I shall not overburden this essay with an attempt to disarm skepticism. But it is worth noting that nearly all currently defended responses to skepticism, other than transcendental ones, agree in denying that perceptual knowledge must be justified by separately insuring that the enabling conditions hold and the skeptic's defeating conditions do not hold.⁶ And since transcendental responses provide at most general guarantees against skepticism, the only tenable responses, which I know of, that attempt to justify particular percep-

⁶ This remark applies to reliabilist theories, Moorean theories that insist on the directness of perception, Quinean theories that attempt to show that the skeptic's doubt is covertly a bad empirical doubt, and Carnapian theories that attempt to show that the skeptic's question is somehow irrelevant to actual empirical claims. The words 'first' and 'separately' are crucial in my formulations. As against some reliabilist views that try to block skepticism by denying closure principles, I think that we can know that no demon is fooling us. But we know this by inferring it from our perceptual knowledge.

Several philosophers have thought that anti-individualism, combined with the view that we are authoritative about what thoughts we think, provides a "transcendental" response to skepticism. Cf. Hilary Putnam, Reason, Truth, and History (New York: Cambridge, 1981). Putnam's argument is criticized by Anthony L. Brueckner, "Brains in a Vat," this JOURNAL, LXXXIII, 3 (March 1986): 148–167. I agree with Brueckner that Putnam's arguments do not do much to undermine skepticism. But Brueckner seems to hold that, if anti-individualism and the authority of self-knowledge are accepted, one would have an antiskeptical argument. He suggests that the assumption of anti-individualism undercuts the assumption of authoritative self-knowledge. I do not accept this suggestion. I believe, however, that there is no easy argument against skepticism from anti-individualism and authoritative self-knowledge. This is a complicated matter best reserved for other occasions.

tual knowledge claims in the face of skepticism take this route. I think

that it is the right route.

I have maintained that perceptual knowledge of physical objects does not presuppose that one has first checked to insure that the background enabling conditions are fulfilled. The same point applies to knowledge of one's own mental events, particularly knowledge of the sort that interested Descartes. Such knowledge consists in a reflexive judgment which involves thinking a first-order thought that the judgment itself is about. The reflexive judgment simply inherits the content of the first-order thought.

Consider the thought, 'I hereby judge that water is a liquid'. What one needs in order to think this thought knowledgeably is to be able to think the first-order, empirical thought (that water is a liquid) and to ascribe it to oneself, simultaneously. Knowing one's thoughts no more requires separate investigation of the conditions that make the

judgment possible than knowing what one perceives.

One knows one's thought to be what it is simply by thinking it while exercising second-order, self-ascriptive powers. One has no "criterion," or test, or procedure for identifying the thought, and one need not exercise comparisons between it and other thoughts in order to know it as the thought one is thinking. Getting the "right" one is simply a matter of thinking the thought in the relevant reflexive way. The fact that we cannot use phenomenological signs or empirical investigation to discriminate our thoughts from other thoughts that we might have been thinking if we had been in a different environment in no way undermines our ability to know what our thoughts are. We "individuate" our thoughts, or discriminate them from others, by thinking those and not the others, selfascriptively. Crudely put, our knowledge of our own thoughts is immediate, not discursive. Our epistemic right rests on this immediacy, as does our epistemic right to perceptual beliefs. For its justification, basic self-knowledge in no way needs supplementation from discursive investigations or comparisons.7

So far I have stressed analogies between basic self-knowledge and perceptual belief. But there are fundamental differences. A requirement that, to know what thoughts we are thinking, we must be able first to discriminate our thoughts from twin thoughts is, in my view, even less plausible than the analogous position with regard to perceptual knowledge.

⁷ I shall not develop the issue of one's epistemic right to one's authoritative self-ascriptions here. It is an extremely complex issue, which deserves separate

Why? In developing an answer to this question, I want to dwell on some fundamental ways in which perceptual knowledge of physical entities differs from the sort of self-knowledge that we have been featuring. We commonly regard perceptual knowledge as objective. For our purposes, there are two relevant notions of objectivity. One has to do with the relation between our perceptions and the physical entities that are their objects. We commonly think that there is no necessary relation between any one person's abilities, actions, thoughts, and perceptions up to and including the time of a particular perception, on one hand, and the natures of those entities which that person perceptually interacts with at that time, on the other. On any given occasion, our perceptions could have been misperceptions. The individual physical item that one perceptually interacts with at any given time is fundamentally independent from any one person's perceptions—and conceptions. The nature of the physical entity could have been different even while one's perceptual states, and other mental states, remained the same.

This fact underlies a normative point about perception. We are subject to certain sorts of possible errors about empirical objects—misperceptions and hallucinations that are "brute." Brute errors do not result from any sort of carelessness, malfunction, or irrationality on our part. A person can be perceptually wrong without there being anything wrong with him. Brute errors depend on the independence of physical objects' natures from how we conceive or perceive them, and on the contingency of our causal relations to them. The possibility of such errors follows from the fact that no matter what one's cognitive state is like (so, no matter how rational or well–functioning one is) one's perceptual states could in individual instances fail to be veridical—if physical circumstances were sufficiently unfortunate.

There is a second sense in which perceptual knowledge is objective. This sense bears on the relation between one person's perceptions of an object and other persons' perceptions of the same object. The idea is that perceptual knowledge, like all other empirical knowledge, is impersonal. Any observer could have been equally well placed to make an observation. Others could have made an observation with the same type of presentation of the scene, if they had been in the same position at the relevant time. And this possible observation could have had the same justificatory status as the original observation. Even though empirical commitments must be made by persons, nothing relevant to the justification of any empirical commitment regarding the physical world has anything essentially to do with any particular person's making the commitment.

The paradigmatic cases of self-knowledge differ from perceptual knowledge in both of these respects. To take the first: in the case of cogito-like judgments, the object, or subject matter, of one's thoughts is not contingently related to the thoughts one thinks about it. The thoughts are self-referential and self-verifying. An error based on a gap between one's thoughts and the subject matter is simply not possible in these cases. When I judge: I am thinking that writing requires concentration, the cognitive content that I am making a judgment about is self-referentially fixed by the judgment itself; and the judgment is self-verifying. There is a range of cases of selfknowledge which extend out from this paradigm. I think that, in all cases of authoritative knowledge, brute mistakes are impossible. All errors in matters where people have special authority about themselves are errors which indicate something wrong with the thinker. Dealing with the whole range requires subtlety. But the point as applied to what I take to be the basic cases is straightforward. No errors at all are possible in strict cogito judgments; they are selfverifying.8

The paradigmatic cases of self-knowledge also differ from perceptual knowledge in that they are essentially personal. The special epistemic status of these cases depends on the judgments' being made simultaneously from and about one's first-person point of view. The point of view and time of the judgment must be the same as that of the thought being judged to occur. When I judge: I am thinking that writing requires concentration, the time of the judgment and that of the thought being judged about are the same; and the identity of the first-person pronouns signals an identity of point of view between the judge and the thought being judged about. In all cases of authoritative self-knowledge, even in those cases which are not "basic" in our sense, it is clear that their first-person character is fundamental to their epistemic status.

These differences between perceptual knowledge and authoritative self-knowledge ground my claim that it is even less plausible than

⁸ Mistakes about the *res* in *de re* judgments are not counterexamples to the claim that basic cogito-like judgments are self-verifying (hence infallible). Suppose I judge: I am thinking that my aunt is charming; and suppose that the person that I am judging to be charming is not my aunt (I have some particular person in mind). It is true that I am making a mistake about the identity of the person thought about; I have no particular authority about that, or even about her existence. But I am not making a mistake about what I am thinking about that person; there is no mistake about the intentional act and intentional content of the act. Authority concerns those aspects of the thought which have intentional (aboutness) properties. For me, those are the only aspects of the content of a thought.

it is in the case of perceptual knowledge to think that basic self-knowledge requires, as a precondition, knowledge of the conditions that make such knowledge possible.

Let us think about the difference as regards objectivity in the relation to an object. In the case of perceptual knowledge, one's perception can be mistaken because some counterfeit has been substituted. It is this possibility which tempts one into the (mistaken) view that, to have perceptual knowledge, one must first know something that rules out the possibility of a counterfeit. But in the cases of the cogito-like self-verifying judgments there is no possibility of counterfeits. No abnormal background condition could substitute some other object in such a way as to create a gap between what we think and what we think about. Basic self-knowledge is self-referential in a way that insures that the object of reference just is the thought being thought. If background conditions are different enough so that there is another object of reference in one's self-referential thinking, they are also different enough so that there is another thought. The person would remain in the same reflexive position with respect to this thought, and would again know, in the authoritative way, what he is thinking.

For example, imagine a case of slow switching between actual home and actual twin-home situations. In the former situation, the person may think "I am thinking that water is a liquid." In the latter situation, the person may think "I am thinking that twater is a liquid." In both cases, the person is right and as fully justified as ever. The fact that the person does not know that a switch has occurred is irrelevant to the truth and justified character of these judgments. Of course, the person may learn about the switches and ask "Was I thinking yesterday about water or twater?"-and not know the answer. Here knowing the answer may sometimes indeed depend on knowing empirical background conditions. But such sophisticated questions about memory require a more complex story. If a person, aware of the fact that switching has occurred, were to ask "Am I now thinking about water or twater?", the answer is obviously "both." Both concepts are used. Given that the thought is fixed and that the person is thinking it self-consciously, no new knowledge about the thought could undermine the self-ascription—or therefore its justification or authority.

In basic self-knowledge, one simultaneously thinks through a first-order thought (that water is a liquid) and thinks about it as one's own. The content of the first-order (contained) thought is fixed by nonindividualistic background conditions. And by its reflexive, self-

referential character, the content of the second-order judgment is logically locked (self-referentially) onto the first-order content which it both contains and takes as its subject matter. Since counterfeit contents logically cannot undermine such self-knowledge, there should be no temptation to think that, in order to have such knowledge, one needs to master its enabling conditions.

The view I constructed on Descartes runs contrary. On that view, since basic self-knowledge is more certain than perceptual knowledge, it is more imperative that one be master of all its enabling conditions. One temptation toward this sort of reasoning may derive from construing self-knowledge as a perfected perceptual knowledge. If one thinks of one's relation to the subject matter of basic self-knowledge on an analogy to one's relation to objects of empirical investigation, then the view that one's thoughts (the subject matter) are dependent for their natures on relations to the environment will make it appear that one's knowledge of one's thoughts cannot be any more direct or certain than one's knowledge of the environment. If one begins by thinking of one's thoughts as objects like physical objects, except that one cannot misperceive or have illusions about them, then to explicate authoritative self-knowledge, one makes one of two moves. Either one adds further capacities for ruling out the possible sources of misperception or illusion in empirical perception, or one postulates objects of knowledge whose very nature is such that they cannot be misconstrued or misconceived. In the first instance, one grants oneself an omniscient faculty for discerning background conditions whose independence from us, in the case of perceptual knowledge, is the source of error. In the second instance, one imagines objects of thought (propositions that can be thought only if they are completely understood, or ideas whose esse is their percipi) whose natures are such that one cannot make any mistakes about them—objects of thought which one can "see" from all sides at once. In either case, one takes oneself to have ultimate insight into the natures of one's thoughts.

This line of reasoning is deeply misconceived. One need only make it explicit to sense its implausibility. The source of our strong epistemic right, our justification, in our basic self–knowledge is not that we know a lot about each thought we know we have. It is not that we can explicate its nature and its enabling conditions. It is that we are in the position of thinking those thoughts in the second-order, self-verifying way. Justification lies not in the having of supplemental background knowledge, but in the character and function of the self-evaluating judgments.

Let us turn to the point that self-knowledge is personal. The view that anti-individualism is incompatible with authoritative self-knowledge is easily engendered by forgetting the essentially first-person character of self-knowledge. We switch back and forth between thinking our thoughts and thinking about ourselves from the point of view of another person who knows more about our environment than we do. This is a key to Descartes's skeptical thought experiments. And it would not be surprising if he tended to think about self-knowledge in such a way as to give it a sort of omniscience from the third-person point of view—in order to protect the first-person point of view from the fallibilities to which impersonal or third-person judgments (especially empirical judgments) are prone. Since we are not omniscient about empirical matters, it is natural to reduce the scope of the relevant third-person perspective so that the character of one's thoughts is independent of an environment about which we cannot be omniscient. Individualism ensues.

To illustrate the train of thought in a more concrete way: we think that we are thinking that water is a liquid. But then, switching to a third-person perspective, we imagine a situation in which the world is not as we currently think it is—a situation, say, in which there is no water for us to interact with. We take up a perspective on ourselves from the outside. Having done this, we are easily but illegitimately seduced into the worry that our original first-person judgment is poorly justified unless it can somehow encompass the third-person perspective, or unless the third-person perspective on empirical matters is irrelevant to the character of the first-person judgment. In this fallen state, we are left with little else but a distorted conception of self-knowledge and a return to individualism.

⁹ My knowledge that I am thinking that mercury is an element depends on an ability to think—not explicate—the thought that mercury is an element. Compare my knowledge that my words 'mercury is an element' are true if and only if mercury is an element. This knowledge depends on understanding the words 'mercury is an element' well enough to say with them, or think with them, that mercury is an element. It is this ability which distinguishes this knowledge from mere knowledge that the disquotation principle as applied to 'mercury is an element' is true (mere knowledge that the sentence "'mercury is an element' is true if and only if mercury is an element" is true). I know that my word 'mercury' applies to mercury (if to anything), not by being able to provide an explication that distinguishes mercury from every conceivable twin mercury, but by being a competent user of the word, whose meaning and reference are grounded in this environment rather than in some environment where the meaning of the word form would be different. The fact that one may not be able to explicate the difference between mercury and every possible twin mercury should not lead one to assimilate one's use of 'mercury' to knowledge of purely formal relationships (e.g., knowledge that all instances of the disquotation principle are true).

As one thinks a thought reflexively, it is an object of reference and knowledge, but simultaneously a constituent of one's point of view. The essential role that the first-person singular plays in the epistemic status of authoritative self-knowledge differentiates this knowledge not only from empirical knowledge, but also from most a priori knowledge, the justification of which does not depend on the first-person point of view in the same way.

The tendency to blur distinctions between a priori knowledge (or equally, knowledge involved in explication of one's concepts) and authoritative self-knowledge is, I think, an instance of Descartes's central mistake: exaggerating the implications of authoritative self-knowledge for impersonal knowledge of necessary truths. One clearly does not have first-person authority about whether one of one's thoughts is to be explicated or individuated in such and such a way. Nor is there any apparent reason to assume that, in general, one must be able to explicate one's thoughts correctly in order to know that one is thinking them.

Thus, I can know that I have arthritis, and know I think I have arthritis, even though I do not have a proper criterion for what arthritis is. It is a truism that to think one's thoughts, and thus to think cogito-like thoughts, one must understand what one is thinking well enough to think it. But it does not follow that such understanding carries with it an ability to explicate correctly one's thoughts or concepts via other thoughts and concepts; nor does it carry an immunity to failures of explication. So one can know what one's thoughts are even while one understands one's thoughts only partially, in the sense that one gives incomplete or mistaken explications of one's thoughts or concepts. One should not assimilate 'knowing what one's thoughts are' in the sense of basic self-knowledge to 'knowing what one's thoughts are' in the sense of being able to explicate them correctly—being able to delineate their constitutive relations to other thoughts.¹⁰

Davidson's views about self-knowledge have some crucial points in common with mine. But he may be making this mistake when he writes that, if one concedes the possibility of partial understanding as I do, one must concede that anti-individu-

One other comparison: I know that I am here (compare: on earth) rather than somewhere else (compare: twin earth). My knowledge amounts to more than knowing I am wherever I am. I have normal ability to perceive and think about my surroundings. I have this knowledge because I perceive my surroundings and not other conceivable surroundings, and I have it even though other places that I could not distinguish by perception or description from here are conceivable. For a variety of reasons, one should not assimilate terms like 'water' to indexicals like 'here'. Cf. "Other Bodies." But these analogies may be helpful here.

For its justification, basic self-knowledge requires only that one think one's thoughts in the self-referential, self-ascriptive manner. It neither requires nor by itself yields a general account of the mental kinds that it specifies. Conceptual explication—knowledge of how one's thought kinds relate to other thought kinds—typically requires more objectification: reasoning from empirical observation or reflection on general principles. It requires a conceptual mastery of the conditions underlying one's thoughts and a conceptual mastery of the rules one is following. These masteries are clearly beyond anything required to think thoughts in the second—order, self-ascriptive way. Explicative knowledge is neither self-verifying nor so closely tied to particular mental events or particular persons' points of view.¹¹

Despite, or better because of, its directness and certainty, basic self-knowledge is limited in its metaphysical implications. It is none-theless epistemically self-reliant. By itself it yields little of metaphysical interest; but its epistemic credentials do not rest on knowledge of general principles, or on investigation of the world.

TYLER BURGE

University of California/Los Angeles

alism undermines the authority of self-knowledge. Cf. his "Knowing One's Own Mind," Proceedings and Addresses of the American Philosophical Association, LX (1987): 448. Cf. also "First Person Authority," Dialectica, XXXVIII, 2–3 (1984): 101–111. It is unclear to me why Davidson says this. I have discussed the distinction between the sort of understanding necessary to think and the sort of understanding necessary to explicate one's thoughts, in "Individualism and the Mental"; "Intellectual Norms and Foundations of Mind"; "Frege on Sense and Linguistic Meaning," forthcoming in The Analytic Tradition, David Bell and Neil Cooper, eds. (New York: Blackwell); and "Wherein is Language Social?" forthcoming in a volume edited by Alexander George (New York: Blackwell).

As I indicated earlier, basic self-knowledge is at most an illuminating paradigm for understanding a significant range of phenomena that count as self-knowledge. Thus, the whole discussion has been carried out under a major simplifying assumption. A full discussion of authoritative self-knowledge must explicate our special authority, or epistemic right, even in numerous cases where our judgments are not self-verifying or immune to error. I think, however, that reflection on the way that errors can occur in such cases gives not the slightest encouragement to the view that anti-individualism (as regards either the physical or social environments) is a threat to the authority of our knowledge of the contents of our thoughts.

REPLY TO BURGE*

YLER Burge observes that there is an apparent conflict between two claims both of which he holds to be true; these are that the contents of our thoughts are individuated in part on the basis of external factors of which the thinker may be ignorant, and that thinkers are authoritative with respect to the contents of their thoughts. He concludes that there is no conflict. The basic reason is that what determines the contents of thoughts also determines what the thinker thinks the contents are. I cannot quarrel with any of this, for like Burge I accept the two claims, have recently discussed the apparent conflict between them, have arrived at the same conclusion as Burge, and have given the same basic reason why there is no conflict. Points of disagreement remain, however, some of them fundamental; I hope the extent of our agreement will help clarify if not resolve our differences.

Burge holds that "basic self-knowledge" is certain; "an error . . . is simply not possible." In support of this dubious thesis he says, "One is thinking that p in the very event of thinking knowledgeably that one is thinking it." But the argument shows only that one cannot go wrong in identifying the content of an attitude, which is not a reason why one cannot go wrong about the existence of the attitude. A more serious worry is that, having described the asymmetry between first— and third—person knowledge of thoughts in such individualistic terms, Burge cannot show that the two sorts of knowledge have the same object. Unless the asymmetry is explained in terms of the relations among speakers, I doubt that a nonskeptical solution is possible.

A related issue arises in connection with Burge's discussion of perceptual knowledge. He says that, though any particular perceptual judgment may be in error, such judgments cannot generally be wrong because, as he puts it elsewhere, perceptual representations "represent what . . . they normally stem from and are applied to." I have long held this view myself (though perhaps unlike Burge I find in it the makings of a cogent argument against some forms of skepti-

¹ In "Knowing One's Own Mind," Proceedings and Addresses of the American Philosophical Association, IX (1987): 441-458.

² "Cartesian Error and the Objectivity of Perception," in Subject, Thought, and Context, Philip Pettit and John McDowell, eds. (New York: Oxford, 1986), p. 131.

0022-362X/88/8511/0664\$00.50

© 1988 The Journal of Philosophy, Inc.

^{*} Abstract of a paper to be presented in an APA symposium on Individuation and Self-knowledge, December 30, 1988, commenting on a paper by Tyler Burge, this JOURNAL, this issue, 649–663.

cism³), but I am not sure why Burge holds it. Part of my trouble is that I do not find Burge's thought experiments as persuasive as he does; they seem to me at best to encourage us to consider what principles we use in (correctly) interpreting the words and thoughts of others. How do we decide what representations normally stem from? We cannot answer "what they represent," since the cause was supposed to answer this very question. But which of the many possible causes is the right one? Events in the nervous system, patterns of stimulation of the nerve endings, or something further out? Burge suggests that we look to the relation between different observers' perceptions of the same object; the idea seems to be that the objects of such perceptions are "impersonal" because different observers similarly placed could have had similar perceptions. But this similarity of content is just what needs to be established.

What is needed in order to give objective content to perceptions, words, or thoughts is not only causal interaction between different observers and the same objects or events, but the right sort of causal interaction between the observers in their shared environment; in a word, communication. Neither knowledge of one's own mind nor knowledge of the "outside" world is possible except in a social setting; "impersonal" thoughts, like other thoughts, depend on interpersonal connections.

DONALD DAVIDSON

University of California/Berkeley

⁸ See "A Coherence Theory of Truth and Knowledge," in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, Ernest LePore, ed. (New York: Blackwell, 1986).

Digitized by Arya Samaj Foundation Chennai and eGangotri

THE HISTORY OF PHILOSOPHY AS A DISCIPLINE*

HE history of philosophy is the subject of a discipline of its own. Its aim is to reconstruct the history of philosophy in such a way that we can understand why philosophy got started in the first place and why it evolved in the way it did, up to and including the present day. For lack of another name, this discipline itself, just like its subject, gets called the "history of philosophy," too. Little thought has been given to the nature of the history of philosophy as a discipline. As a result, there is a good amount of confusion, not only as to what historians of philosophy try to do, but also as to how they ought to go about doing it. It would even seem that some of the historians' own work reflects such confusion. Hence, it seems appropriate to try to clarify, as well as we can, what a historian of philosophy is attempting to do.

Part of the confusion seems to be due to a misleading ambiguity in the term 'history of philosophy'. Historically it has been used in two rather different ways, each of which corresponds to a very different tradition of treating the history of philosophy, both of which persist to the present day but tend to get conflated.

From roughly the middle of the seventeenth century onwards, we find treatises with the title "History of Philosophy." Perhaps the earliest of these is Georg Horn's Historia Philosophica (Leiden, 1655); the most famous clearly is Jacob Brucker's Historia Critica Philosophiae (Leipzig, 1742–1767). If we look at these treatises, we are surprised to find that they are not histories in our sense at all; they do not try to trace the development of philosophy from its beginnings; they do not even follow the chronological order. They show themselves to stand in a much older tradition that goes back to antiquity, namely, the doxographical tradition. Almost from the beginning, certainly from Aristotle onward, there have been philosophers who have studied the history of philosophy for philosophical reasons. They were interested in philosophical views or positions of the past, because they thought that at least some of them were still worth philosophical consideration, perhaps even true in some im-

0022-362X/88/8511/0666\$00.70

© 1988 The Journal of Philosophy, Inc.

^{*}To be presented in an APA symposium on the History of Philosophy as a Discipline," December 28, 1988. Alan Donagan will be co-symposiast, but his paper was not received in time to be published in this issue. Kenneth Schmitz will comment; see this JOURNAL, this issue, 673/4.

portant regard, but perhaps also just false, yet false in an interesting, revealing, paradigmatic way. There was also the widespread philosophical conviction that there is a basic set of philosophical questions and that it might be philosophically worthwhile to scan the history of philosophy systematically for different answers to these questions. It was one or another version of this assumption, usually some form of eclecticism, which inspired the earlier large-scale systematic treatments of the history of philosophy, e.g., Diogenes Laertius's *Lives* in antiquity, and, with it as a model, the early modern treatments which, from some point in the seventeenth century onward, came to be called "histories of philosophy." When at the end of the eighteenth century Kant talks about historians of philosophy, what he has in mind still are philosophical doxographers of this kind, rather than historians in the sense I am interested in.

But toward the end of the eighteenth century, a very different tradition emerges. Meiners's history of 1786 seems to be the first to adopt a chronological disposition, and in the next decade appear, in rapid sequence, the histories of Tiedemann (1791 ff.), Buhle (1796), and Tennemann (1798 ff.), which make it their aim to trace the development of the history of philosophy from its beginnings to the present day.

As opposed to their doxographical predecessors, these histories originally are written out of the conviction that the philosophical positions of the past are no longer worth considering philosophically, that they are out of date; if they are still worth considering at all, it is because they constitute the steps through which we historically arrived at our present philosophical position. Thus, they are still histories written from a philosophical point of view, in fact from a particular philosophical position; and they regard the past, the history of philosophy, as leading up to this position. They are sometimes written to show how the given position is the result of a long historical process in the course of which we have come nearer and nearer to the truth.

But it is easy to see that the enterprise of reconstructing the development of philosophy, though originally inspired by such philosophical convictions and interests, in fact does not rest on them. And so, in the course of the nineteenth century, we see how these philosophical assumptions about the history of philosophy get shelved by historians like Eduard Zeller. What emerges is a discipline that, with the tools of the historian, tries to do no more, but also no less, than to reconstruct historically the development of philosophy. It does not

make any assumptions as to whether the views of the past are still worth considering philosophically or not; it certainly does not proceed selecting those views or positions of the past which might be thought to be of continued philosophical interest; it does not itself take a philosophical position and tries to reconstruct the past as leading up to it; it does not see itself at all as serving the interests of philosophy, or any other discipline, for that matter. It is history of philosophy in this sense with which I am concerned, rather than with the very different enterprise of history of philosophy in the philo-

sophical, doxographical tradition.

I have no objection to a philosophically oriented study of the history of philosophy in the doxographical tradition, though I find that the use of the word 'history' for this sort of study is somewhat misleading. If I insist on the distinction it is because it is often overlooked, especially by philosophers, though there is a fundamental difference, both in principle and in practice, and because I think that the kind of history of philosophy in the doxographical tradition which philosophers continue to practice to the present day, a study which imposes our philosophical views and interests on the history of philosophy, ultimately presupposes the second kind of history of philosophy, i.e., a study of the history of philosophy in its own right, on its own terms, quite independently of our philosophical views, interests, and standards. And this for the following reasons: it had always been, in fact, though not in principle, a weakness of the doxographical tradition to underrate the enormous difficulties involved in precisely identifying a view of the past, especially of the more distant past, and in representing it in such a way as to make it accessible to philosophical consideration in terms of the contemporary debate and to comparison with other contemporary views. Once we become aware of the enormous difficulty, we have to make a choice. We can choose, perfectly legitimately, to forego the enormous difficulties involved in identifying a view, say, of Aristotle's, by settling for a view which, if not Aristotle's itself, seems to be a view very much like it and, in any case, is a view of philosophical interest. But, equally legitimately, we may choose not to compromise and to insist on identifying Aristotle's view as well as this is possible. But, if we do opt for the latter, I think we have to study the history of philosophy on its own terms. For we will only be able to identify confidently a view of the past, if we have a thorough understanding of the historical context in which it was held, an understanding of which views were available in this context and which not. And we will not have this kind of grasp on the immediate context, unless we have a solid grasp on a fairly large context. And this larger context inevitaDigitized by Arya Samaj Foundation Chennai and eGangotri bly will be crucially characterized by many details that are of no concern to the philosopher, either because they are not philosophical in the first place, or because they are of no philosophical interest to us. After all, most of the history of philosophy is of no or little philosophical interest to us. Hence, the only way, it seems to me, to get a sufficiently adequate grasp on the historical context in which a view was held, to allow us to identify a view with sufficient reliability, is to study the history of philosophy in its own right, including all those details which are of little or no philosophical interest or even unphilosophical in character, rather than to study selectively just those parts of the history which to us still seem to be of philosophical interest. So, even if in the end we are interested in the history of philosophy for philosophical reasons, we may decide that these interests are best served by a study of the history which is independent of these interests. After all, we may come to believe that Aristotle's or Thomas's actual views are likely to be a lot more interesting philosophically than their doxographical substitutes. Once we have such a historical study we are in a much better position to judge whether philosophical positions of the past continue to be of philosophical interest or not.

Perhaps this notion of the history of philosophy as a discipline in its own right will become clearer if we look at how a historian of philosophy, as opposed to a philosopher interested in the history of philosophy for philosophical reasons, will treat a view of the past. Let us assume that the view, along with the reasons given for it, has been correctly identified. Now, the philosopher will be interested in the view and the reasons given for it as such. He will ask questions like: Is the view true, reasonable, plausible, possible, or not? Are the reasons offered for it adequate or even conclusive? Which other reasons could one advance in favor of the view, which ones do speak against it? The historian, on the other hand, is interested, not in the view and the reasons as such, but in the historical fact that a certain person in a certain historical context held this view and gave these reasons for it. The questions he will ask are not whether the view is true or the reasons are adequate, but rather whether the view would have seemed to be true or plausible at the time, whether at that point in the past the reasons offered would have been taken to be adequate or conclusive.

To put the matter differently: if we try to understand why a contemporary of ours holds a certain view, we look at the reasons he offers for this view. Depending on whether these reasons seem good reasons or not, we think we have understood why he holds this view. If the reasons are perfectly good reasons, we will be inclined to think

Digitized by Arya Samaj Foundation Chennai and eGangotri that he holds the view for these reasons; if not, we know that we have to look for a more complicated explanation of why he holds the view. But, if we consider somebody in a very different historical context. the situation is rather different. What matters in that case is whether the reasons offered would have counted as good reasons at the time or not. For though the reasons he offers, by our light, are excellent reasons, it may turn out that at the time they could not reasonably have been regarded as good reasons. So we cannot explain his belief just in terms of these reasons, though they seem adequate to us. Or suppose the reasons offered are bad reasons, so bad that nobody nowadays would hold the view on the basis of them; hence, if nowadays somebody gave them as his reasons for the view, we would be inclined to think that the explanation for his holding the view must lie somewhere else. But it might very well be the case that, in the historical context we are considering, this line of reasoning would have appeared perfectly adequate and persuasive. So the historian will explain somebody's holding a view in the past not with reference. e.g., to what would seem to us to be a perfectly adequate line of argument, but to a line of argument which would have seemed perfectly adequate then. But this means that he now has to go on to explain why this line of reasoning would have seemed perfectly adequate then, if we are to understand why the philosopher in question held this view for these reasons. So the historian does try to understand and to explain, as far as he can, the philosophical views of the past in terms of philosophical arguments and philosophical considerations. But they are the philosophical considerations of that time, rather than ours. In this sense the historian explains the facts of his history out of their historical context.

It might be worthwhile to look at this a bit more closely. Within the relevant historical context, we have to distinguish between the narrower context, which is roughly something like the philosophy of the time, and the larger historical context, which is the rest of the culture of the time, which includes the social, political, economical, and religious conditions and whatever else may be relevant, for instance, the state of the sciences. We have to make this distinction for various reasons. One is the following: as we said earlier, the reasons advanced for a view may be so inadequate, even in terms of the philosophy of the time, that we come to think that the view was not held for this reason, but for some extra-philosophical reason. In this case the historian has to explain the fact that the view was held in terms of the larger historical context.

But the distinction is also important for the following reason.

Philosophy is an ongoing enterprise with its own standards of what is acceptable and what not, with its own standards, e.g., as to which line of reasoning is acceptable in support of a view, as to what counts as a reasonable view. Perhaps these standards change over time, as philosophy develops on the basis of purely philosophical considerations. Perhaps it also is the case that these standards change over time under the pressure of the larger context, e.g., the attraction science exercises. But, however much the philosophy of a time may be embedded in the rest of the culture of the time, at each time these standards and rules of the subject itself offer it a considerable amount of autonomy from the rest of the culture, though this autonomy may be larger or smaller at different times.

Now, at least in our tradition, it is part of the enterprise of philosophy to hold views for reasons that are adequate by the standards of the enterprise. And the autonomy of the subject is reflected precisely by the fact that, by and large, if philosophers do claim to hold a certain view for certain reasons, they actually do hold it for these reasons, because they assume these reasons to be adequate by the standards of the enterprise. Hence, however much other explanations in terms of the larger context may be available, it is the task of the historian of philosophy to explain philosophical views or positions, as far as possible, in terms of the philosophical considerations that were taken to support them. Perhaps we may call a history that follows this principle an "internal" history of philosophy, since it explains the development of philosophy, as far as possible, in terms of considerations internal to philosophy. But even such an internal history will have to rely, again and again, on factors other than philosophical considerations. In fact, given how embedded philosophy is in the rest of the culture, both at a time and through time, one also could write an external history of philosophy. Its lack of detail when it comes to particular philosophical views or positions would reveal the degree of autonomy philosophy has enjoyed.

We now have a very schematic idea of how a historian, out of the historical context in which a view was held, explains the fact that somebody held his view. But we can also faintly see in terms of this schema how the historian might explain the development of philosophy. For suppose the view in question is a new view, a relatively central view, and the reasons given for it at the time seem plausible, so that the view finds considerable acceptance. As a result the narrower historical context will change more or less significantly. And, as a result of this, the kinds of philosophical considerations which are

regarded as acceptable will change more or less significantly. So what the historian will look for are not new views that we find philosophically interesting, but new views whose appearance significantly changed the context. By explaining how they came to be held and how they affected the context, we will slowly get a picture of the development of philosophy which in no way depends on our philosophical views and interests. In fact, these themselves will be included in the picture. For the philosophical views of the present should allow for exactly the same kind of explanation as the views of the more recent or the more distant past, namely, in terms of a narrower and a larger historical context. And the narrower context will be explained, as usual, in terms of a change in context, due, e.g., to the appearance of new views or positions, against the background of which a view appears eminently reasonable.

That even the philosophy of the present should be accounted for in this way by the historian may strike us as curious, but it is just the result of assuming that the historian does not approach the history of philosophy as a philosopher with his own philosophical views and interests. The philosopher, of course, explains his views by explaining why they are true, and not by explaining why, given his historical context, he has them or takes them to be true.

MICHAEL FREDE

Princeton University

THE HISTORY OF PHILOSOPHY AS ACTUAL PHILOSOPHY*

s Michael Frede has indicated, our appreciation of the role played by the history of philosophy in the career of philosophy itself has grown. To speak of it as a discipline points up two aspects of that relationship. First, there is the claim to a place among the historical disciplines, since the history of philosophy has its own subject matter more or less defined by the character of philosophy and yet it appropriates the general methods of historical research and modern historiography. There is no doubt that the development of the latter has transformed the standards of historical research and the writing of history. Secondly, we are better situated today to recognize the historical character of actual philosophy itself. To be sure, it is the study of reality in terms of ideas and judgments of a rational kind, composed of concepts and with criteria of evidence and canons of argumentation. Nevertheless, it is a human enterprise that has developed in and through time, with its own traditions of inquiry, its own "canon" of writings (always subject to revision), and its own internal pluralism. Despite some appearances, philosophy does not reinvent itself with each generation, even though some philosophers may claim to do so.

If we are to appreciate the distinctive role of the history of philosophy, we must attend to its actual, concrete character. For there is an important sense in which the written history of philosophy records the career of what is both a tradition and a community. Now, a tradition is in a way a community built up in and through time. Aristotle remarked of artisans that the development of their skills required the handing on of what had already been acquired by past generations. This traditio is the condition for both the preservation of tried techniques and the possibility for new or improved techniques. We must not press the analogy too closely, however, lest we suppress the original character of philosophical inquiry. It does sometimes build upon earlier gains, but it also returns to the beginning by calling into question the very starting point of inquiry. But because—or so it seems to me—the starting point of philosophical inquiry is richly concrete, there are many ways of grasping it, even as there are many objectives pursued by inquiry. And so, there results an inherent pluralism in philosophy. This need not condemn philo-

0022-362X/88/8511/0673\$00.50

© 1988 The Journal of Philosophy, Inc.

^{*}Abstract of a paper to be presented in an APA symposium on the History of Philosophy as a Discipline, December 28, 1988, commenting on a paper by Michael Frede, this JOURNAL, this issue, 666–672.

sophy to relativism or to self-contradiction. Indeed, there are fewer contradictions and fewer disagreements in philosophy than we might at first think. The internal problem of communication between philosophers is rather the opposite: to find enough common ground on which to disagree in a strong and genuine way. Perhaps the endemic emergence of schools is prompted by the search for sufficient common ground on which differences can be fruitfully discussed.

But, of course, the history of philosophy runs more deeply and broadly than any school or movement. By 'history of philosophy' I do not mean the history of ideas, at least as it is sometimes understood. For there is a view—admittedly less prevalent now that hermeneutics has spoken its word—that treats ideas as units of meaning, relatively independent of one another, but whose movement can be traced from one thinker to another, and which can be found in various particular compositions in individual thinkers. I should prefer to call such a view a sort of mechanics of ideas rather than the history of philosophy. For the history of philosophy starts from a more concrete base, and its integers are neither facts nor ideas, but persons. Ideas do not live a free life of their own, but are taken up rather into the personal thought of a philosopher and suffused with the energy of his own mind and personality. It goes without saying that conclusions must be drawn in accordance with the canons of logical reasoning, and are subject to criticism by those same canons. Nevertheless, more needs to be said, for hidden possibilities within ideas and new interrelationships among them are disclosed in the medium of actual philosophical discourse. May we even say more? Are not the very ideas themselves transformed by the energy of a great philosophical mind? For ideas have existence only in minds and minds only in persons. The only being ideas have is the being of the philosopher who thinks them. Moreover, the very actuality of thought places them in relationships that could hardly be drawn forth by the formal canons of logical analysis alone. There is, if I may speak metaphorically, a sort of concrete, existential "chemistry" in which received ideas take on a new life and shape as they enter into unprecedented relationships, being transformed by the light energy of actual philosophical insight. The logical justification of such new intelligible shapes and affinities may well be recognized after a philosopher has shown them to us, but it is then after the fact that we look back to the logico-rational basis. That is why philosophy is a life, and why it has a history and not simply a formal logic.

KENNETH L. SCHMITZ

University of Toronto

New Forthcoming from Hackett

NELSON GOODMAN & CATHERINE Z. ELGIN

Reconceptions in Philosophy & Other Arts & Sciences

188 pp. \$21.50 prof. price \$17.50 8 pp. color plates.

"The idea that knowledge and truth are not as absolutely crucial as we have thought is absolutely crucial; it is right."

—ALEXANDER NEHAMAS

I Knowing & Making II How Buildings Mean III Interpretation & Identity: Can the Work Survive the World? IV Variations on Variation—or Picasso back to Bach V Sights Unseen VI Inertia & Invention VII Confronting Novelty VIII Representation Represented IX The Epistemic Efficacy of Stupidity X A Reconception of Philosophy

New HPC Classics:

MILL, The Subjection of Women Edited by Susan M. Okin

128 pp. \$4.95 exam copy \$1.50 "... beautifully produced. Clear and helpful introduction by one of the leading feminist scholars of our generation. It will be invaluable for teaching and scholarship alike."

-IAN SHAPIRO, Yale University

HEGEL, Introduction to The Philosophy of History

with selections from

The Philosophy of Right

New translation by Leo Rauch

160 pp. \$4.95 exam copy \$1.50
"...as felicitous as a translation of Hegel can be."

—HENRY S. HARRIS

Forthcoming Spring '89

PLATO, Symposium

New translation by P. Woodruff & A. Nehamas

(t) \$3.45

exam copy \$1.00

A single edition of **Donald A. Cress's** highly acclaimed translation of

ROUSSEAU, Social Contract

112 pp. \$2.95 exam copy \$1.00 Booths 104, CC-0. In Public Domain, Gurukul Kangri Collection, Haridwa

IONATHAN BENNETT

Events & Their Names

274 pp. \$16.50 prof. price \$12.50

"Blows away many pockets of unclarity that still exist in the current discussion of event and causation, and raises the debate ... to a new level of illumination and precision."

—JAEGWON KIM

The Scientific Background

to Modern Philosophy

Edited by Michael R. Matthews

ca. 144 pp. \$17.50 cloth \$4.95 paper spring 1989 exam copy \$2.00

A supplementary text for courses in the history of modern philosophy, helping to link developments in modern science and modern philosophy.

Aristotle • Copernicus • Bacon • Galileo Descartes • Boyle • Huygens • Newton

ARISTOTLE

Prior Analytics

New translation with Commentary by Robin Smith

336 pp. \$27.50 cloth prof. price \$17.50 This translation is the first new English version since Tredennick's of 1938 and is the only English translation with commentary. The exposition is clear and straightforward, reflecting a deep knowledge of the text.

Three Distinguished Works Brought Back into Print:

JONATHAN BENNETT

Rationality

(t) \$17.50 cloth spring 1989 (t) \$6.50 paper prof. price \$5.00

W.D. ROSS

The Right & the Good

186 pp. \$14.50 prof. price \$10.50

HENRY SIDGWICK

Outlines of the History of Ethics Reprinting of the 5th Edition of 1902.

320 pp. \$7.95 prof. price \$3.50

Hackett

Publishing Company
Box 44937, Indianapolis, IN 46204

Booths 104, 106, 108 APA

Digitized by Arya Samaj Foundation Chennai and eGangotri
A New Critical Edition from the Complete
Works of George Santayana

The Sense of Beauty

George Santayana Herman J. Saatkamp, Jr., General Editor William G. Holzberger, Textual Editor Introduction by Arthur C. Danto



"Beauty is pleasure regarded as the quality of a thing" states George Santayana in this acclaimed book. Grounded on a naturalistic look at human psychology, *The Sense of Beauty* is Santayana's insightful approach to the study of aesthetics. Even today, it is one of the most widely read volumes in all of Santayana's vast philosophical work. *The Sense of Beauty* is the second volume in a critical edition of the complete works of George Santayana, a series which will be a "new starting point for exploring the connection between Santayana's life and ideas." — *The New York Times Book Review*Illustrated \$35.00

Other volumes in the series will include:

- · Persons and Places (now available)
- · Interpretations of Poetry and Religion
- · The Last Puritan
- · The Life of Reason
- Three Philosophical Poets: Lucretius, Dante, and Goethe
- Winds of Doctrine: Studies in Contemporary Opinion
- · Egotism in German Philosophy
- · Character and Opinion in the United States
- Soliloquies in England and Later Soliloquies
- · Scepticism and Animal Faith
- Dialogues in Limbo
- · Platonism and the Spiritual Life
- · Realms of Being
- The Genteel Tradition at Bay and Some Turns of Thought in Modern Philosophy
- · The Idea of Christ in the Gospels
- · Dominations and Powers
- Collected Poems
- · Collected Articles, Reviews, Essays, and Marginalia

The MIT Press

55 Hayward Street • Cambridge, MA 02142

CC 0. In Public Domain, Gurukul Kangri Collection, Haridwar

THE JOURNAL OF PHILOSOPHY

VOLUME LXXXV, NO. 12, DECEMBER 1988

VIRTUE CONCEPTS AND ETHICAL REALISM

N EVERAL philosophers have recently suggested that certain concrete ethical concepts, such as those of promising, courage, kindness, and dishonesty, provide rich and finely-grained evaluations.1 They also hold that such concepts refer to properties about which some form of "ethical realism" is true.

These philosophers assume that the concepts in question are evaluative or have an evaluative component. I disagree. I shall argue that many of these concepts are merely descriptive and do not express an evaluation. I call this position descriptivism, even though this term

* I would like to thank Susan F. Brower, Joseph L. Camp, Jr., Walter Edelberg, and Brad Hooker for their insightful comments on an earlier draft of this paper.

¹ This sort of view is endorsed by Sabina Lovibund, in Realism and Imagination in Ethics (Minneapolis: Minnesota UP, 1983); John McDowell, in "Are Moral Requirements Hypothetical Imperatives?," Proceedings of the Aristotelian Society, LII, Supplementary Volume (1978): 13-29; "Virtue and Reason," The Monist, LXII, 3 (July, 1979): 331-350; "Non-cognitivism and Rule-following," in Steven H. Holtzman and Christopher M. Leich, eds., Wittgenstein: To Follow a Rule (London: Routledge & Kegan Paul, 1981), pp. 141-162; and "Values and Secondary Qualities," in Ted Honderich, ed., Morality and Objectivity (London: Routledge & Kegan Paul, 1985), pp 110-129; Mark Platts, in Ways of Meaning (London: Routledge & Kegan Paul, 1979), pp. 243-262; and David Wiggins, in "Truth, Invention and the Meaning of Life," Proceedings of the British Academy, LXII (1976): 331-378.

Lovibund holds that 'brave', 'malicious', and 'corrupt' are used in "lower-level evaluations" (pp. 14/5). In the first three articles mentioned, McDowell concentrates on 'courage' and 'kindness'. Platts claims to be an ethical realist about language containing the expressions 'sincerity', 'loyalty', 'honesty', 'prudence', 'courage', and 'integrity' (pp. 243/4). Wiggins mentions 'brave', 'dishonest', 'ignoble', 'just', 'malicious', and 'priggish' (p. 359).

Bernard Williams, in Ethics and the Limits of Philosophy (Cambridge: Harvard, 1985), also endorses this view. He argues, however, that the properties referred to by concrete ethical evaluations have no place in the world described by the "absolute conception of reality." Yet he holds that concrete ethical concepts are both evaluative and descriptive, and that judgments using these concepts can be true and known.

0022-362X/88/8512/0675\$01.90

© 1988 The Journal of Philosophy, Inc.

has been overworked. I use antidescriptivism to refer to the opposing view. Descriptivists can be realists about the properties referred to by these concepts, but they are not thereby ethical realists. So a descriptivist can be a realist about courage, for instance, without being an ethical realist.

I shall follow a number of writers by concentrating on concepts of what have been taken to be virtues and vices. As it turns out, strong arguments can be made that certain traditional virtue and vice concepts are evaluative. I hope to show that such arguments are unsound. I shall argue that, for any traditional virtue, we can imagine worlds at which the virtue is of negative value overall. So, for any virtue, we can imagine a world where having that virtue is not good, and acting in accord with that virtue is not right. This conflicts with the most interesting form of antidescriptivism.

Descriptivism can be illustrated by considering a concept that is not of a virtue or a vice. Williams claims that the concept of a lie has an evaluative component (op. cit., p. 140). According to his view, to say that someone lies is not only to say that she states a falsehood with the intent to deceive, but it is also to express a negative evaluation of her or her action. In contrast, the descriptivist holds that to say that someone lies is merely to describe, not to evaluate.

Without at this point arguing for descriptivism, let me compare the concept of lying with the more obviously descriptive concept of kicking other people. Several things are worth noting. First, descriptivism grants that lying is usually of negative value; but the value is not included in the concept of lying. Similarly, kicking others is usually of negative value; but the concept of kicking others is not evaluative. Second, speakers who discuss someone's lying often expect their listeners to form negative evaluations. The descriptivist holds that this is because the speakers take their listeners to have values close to their own, not because the concept of lying is evaluative. Similarly, if I tell you that someone kicks others, I merely describe; but I assume that you will form a negative evaluation of the other person. Third, there are several senses in which the descriptivist can be a realist about lying, for lying exists, and facts about lying can be true independent of our beliefs. Similarly, one can be a realist about the property of kicking others. Thus, descriptivism is not committed to "antirealism" or "nonrealism" about the properties referred to by concrete ethical concepts.2 Fourth, it would be easy to

² The expression 'concrete ethical concepts' may suggest concepts that do express an ethical evaluation. As used here, however, it is intended to refer to concepts which are relatively concrete and which have traditionally been thought to express ethical evaluations. Concepts may be said to be more or less concrete depending on

give an account of evaluation—say, in terms of emotional responses—from which antidescriptivism immediately follows. So there are "cheap" forms of antidescriptivism according to which both the concept of lying and the concept of kicking others have negative value. But these forms of antidescriptivism would hold true of any concept that refers to properties that, as a matter of fact, we value or do not value. The form of antidescriptivism I wish to consider is more interesting, because it does not easily apply to a broad range of concepts.

The term 'descriptivism' could mislead, since its use may seem to presuppose noncognitivism. Traditional noncognitivists distinguish between descriptive propositions, which are held to apply to the world independently of our attitudes, and evaluations, which are held to express our attitudes, and are viewed as inapplicable to the world that is independent of our attitudes. The usual noncognitivist view is that descriptive propositions, but not evaluations, can be true or false. In this essay, the distinction between description and evaluation is not meant to imply that evaluations cannot be true or false. So the claim that a concept is evaluative is not meant to imply that propositions containing the concept cannot be true or false. One may use 'descriptive' for all propositions that can be true or false. Within this broadly descriptive set of propositions, one may distinguish between the evaluative and the nonevaluative. This intuitive distinction is worth preserving, even if one wishes to avoid noncognitivism

I.

Let us look at concepts as having conditions of application, such that a concept applies only to objects that meet its conditions. Totally competent speakers must recognize that each condition of application is a necessary condition for applying the concept to an object.³

how much determinate descriptive content they have, though this notion must be left intuitive. The concepts of good and right are highly abstract, the concepts of courage and brutality are relatively concrete. I concentrate on concepts of traditional virtues, but I hope that my argument lends support to a skeptical attitude toward antidescriptivism about all concrete concepts.

In the last twenty years, philosophers have argued that speakers can acquire the use of a term even if they are wrong about conditions for its application, or even if the term is best looked at as not having any conditions of application. (Cf. Saul Kripke, Naming and Necessity (Cambridge: Harvard, 1982); and Hilary Putnam, "The Meaning of 'meaning'," in Mind, Language and Reality (New York: Cambridge, 1975), pp. 215–271.) I find this view attractive, so I discuss the requirements on a speaker who is "totally competent" with a "concept" (not a term). In any case, it is not clear that concepts of character traits should be treated in the same fashion as proper names and "natural kind" terms. Moreover, the phrase 'condition of application' is used by Williams and others who discuss concrete ethical concepts. I am simply retaining their usage.

If a concept has a condition of application, then, in any possible world, objects in the extension of the concept must meet the condition. Descriptivism holds that concrete ethical concepts have only descriptive conditions of application. Antidescriptivism holds that concrete ethical concepts have both descriptive and evaluative con-

ditions of application.

Before examining the evaluative condition, consider some examples. One sort of example is of cases in which a concept that we might take to be evaluative is applied to actions from which our usual evaluation is withheld. Examples of this kind were discussed by Philippa Foot⁴ and R. M. Hare.⁵ Foot argued that 'rude' is both descriptive and evaluative. Hare responded that 'rude' is primarily descriptive. He supported this view by citing actions which meet the descriptive conditions for rudeness but which we would not evaluate adversely. Other examples of this sort are easy to imagine. For instance, most of us allow that some actions are honest but wrong, or dishonest but right. Yet such cases show only that there are isolated circumstances in which the concept applies but our normal evaluation is withheld. They are consistent with the view that rudeness is usually of negative value, or the view that honesty is usually of positive value. So, by slightly weakening the position, an antidescriptivist can argue that a concept has an evaluative condition just in case that condition must be met by most actions in the extension of the concept.

A second sort of example involves concepts that refer to properties that are no longer evaluated as they were in the past. For many, chastity is a good example of such a property. Another is pride, which is now often seen as a sign of appropriate self-esteem-especially if the pride is directed at a property that was wrongly denigrated in the past, such as being black. In taking note of these, the antidescriptivist can claim that whether a concept is evaluative is relative to a "form of life," "tradition," "point of view," or "conceptual scheme." I shall say below why I take this move to be inadequate. But the antidescriptivist can also claim that, while some concepts undergo shifts in their evaluative component, such shifts are impossible for central virtue concepts, such as prudence, courage, and kindness.

^{4 &}quot;Moral Beliefs," in Virtues and Vices (Berkeley: California UP, 1978), pp. 110-131.

⁵ Moral Thinking (New York: Oxford, 1981), pp. 17/8, 74/5. Hare argues briefly for a form of what I call descriptivism, but his own use of 'descriptivism' is different than mine.

What we need, then, are examples involving central virtue concepts, such that the virtues in question are shown not merely to have negative value occasionally, but rather to be of negative value in general. Consider these "possible worlds":

World W1: Human beings are enslaved by a physically and intellectually superior species, and are forced to spend their lives constructing spaceships. They have no reasonable hope of successful revolution against this system.

Members of the master species, who are completely informed about humans, regularly review the behavior of each human. If a human overcame fear, ignored fear, or failed to have fear in a situation where most humans would act out of fear, then that human is tortured for the rest of her life. If any other human benefited from the overcoming of fear, the ignoring of fear, or the failure to have fear, then that other human is also tortured. It follows that, if a human acts courageously or benefits from another's courage, she will be tortured for the rest of her life.⁶

If a human does not act courageously, she is allowed to live a moderately fulfilling life.

World W2: As a matter of psychological fact, kind people become weak and powerless. They are indecisive, lose control of their lives, and become ineffective in relations with others. Only those who are not kind can become strong and take charge of their lives. They need not be cruel or vicious, but even isolated kind acts lead to the destruction of one's character.

According to descriptivism, courage has no positive value at W1, and kindness has no positive value at W2. So the concepts of courage and kindness seem to be purely descriptive concepts that apply at other possible worlds to actions that lack positive value.⁷

The descriptivist agrees that courage and kindness have positive value at our world, but holds that our concepts of them are not evaluative. The mere fact that a property is valuable at our world does not imply that a concept that refers to it is evaluative.

One response to descriptivism is that the concept of courage (for

⁷ In this paper, I do not always distinguish between virtue concepts as they apply to actions and virtue concepts as they apply to agents. The context makes clear which application is intended.

⁶ I am not attempting to give a complete account of courage. I do, however, assume that an action is courageous only if the agent overcomes fear, ignores fear, or is fearless in a situation where most people would act out of fear. I express these three in a shorthand manner by saying only that the agent "overcomes fear." My view of courage has been influenced in part by Peter T. Geach, *The Virtues* (New York: Cambridge, 1977); and James D. Wallace, *Virtues and Vices* (Ithaca: Cornell, 1978).

example) only requires that courage be valuable at our world. On this view, a speaker understands the concept of courage if she understands that (a) at other worlds, certain acts meet a given descriptive condition, but (b) at our world, a courageous act must both meet the descriptive condition and have positive value. An analogy is provided by revising our concept of gold. Let gold* be a concept just like our concept of gold, except that in order to apply gold* correctly, a speaker must recognize that members of its extension have positive value at our world. At some other world W*, where gold is abundant, iron is scarce, and socio–economic conditions are similar to ours, gold is not valuable. Suppose, however, that as the concept of gold* is understood, it may be applied to gold objects at W*, even though we recognize that gold (and hence gold*) is not valuable at W*.

A similar response is that the evaluative condition varies with the world, such that, if one is considering a world at which the property denoted by the concept is of a certain positive or negative value, then one must recognize that value of the property, at that world. So actions at other worlds may be described by such concepts, yet at these other worlds the property denoted by the concept has a different value than it has at our world. Nevertheless, at the actual world, any competent speaker must recognize the actual value of the property. Thus, if one fully understands the concept of courage, then to describe an action at our world as courageous is thereby to acknowledge the positive value of the action; but one must also recognize certain actions performed at W1 as courageous and yet see that they are not of positive value at W1.

These responses seem ad hoc, and I do not believe they capture the central antidescriptivist intuitions. But there is a further problem: W1 and W2 are not merely possible worlds. Our world might be like them. In the case of W1, this must occur in the future; but, in the case of W2, we could in principle discover that our current psychological beliefs are wrong, and that our world is at present similar to W2. The descriptivist holds that whether courage and kindness have value depends on how the world is, and that therefore we could discover that they lack value at our world. W1 and W2 are possible states of our world.

Of course, we do not believe our world is like W1 or W2. In daily life we automatically value courage and kindness, because given our basic beliefs about society and psychology (beliefs that we rarely have occasion to doubt), we hold that courage and kindness are valuable. Yet we could find out that these beliefs are false, in which case we would no longer automatically value courage and kindness.

II.

Antidescriptivism can be made more defensible by spelling out the evaluative condition of application along certain lines. Most antidescriptivists have cashed out evaluation in terms of there being a reason for action. The clearest account is developed by Williams, on the basis of work by McDowell. Williams holds that concrete ethical concepts are "thick," where this means that they are "both worldguided and action-guiding" (op. cit., pp. 129-130, 140-142). His idea is that the correct application of such concepts depends both on how the world is and on whether they are understood as "characteristically" or "often" providing a reason for action. He holds that it is a plausible "demand on the understanding" that the user of such concepts must accept their evaluative point (op. cit., p. 142). Let us say that a concept is positive if it denotes a property that we ordinarily take to be of positive value. On Williams's view, a positive, thick concept is action-guiding in that it is correctly applied to an action only if there is a reason for the agent to take that action. A competent speaker misapplies a concept if she applies it to an action for which there is no reason. So, if she knows that the concept applies to an action she is considering, she thereby knows that there is a reason for the action. A negative, thick concept, in contrast, is action-guiding in that it is correctly applied to an action only if there is a reason for not taking the action. If a competent speaker applies the concept to an action, she thereby knows that there is a reason for not taking that action. In general, the evaluative condition of application is provided by this action-guiding element.8

Let us interpret the requirement that there be a reason in a strong fashion. Accordingly, to say that there is a reason for an action is not merely to say that there is some reason, or that the agent has some reason, for the action. Rather, there is a reason for an action only if there is reason to take the action, all things considered. In this sense, there may be reason for or against an action even if the agent does not recognize the reason.⁹

By construing evaluation in terms of there being reasons, the antidescriptivist can respond to descriptivism. The antidescriptivist will

⁹ Williams sometimes formulates his view in terms of prima facie reasons. I com-

ment on prima facie reasons in section IV below.

⁸ This account must be extended in order to apply virtue concepts to agents. A concept of a character trait has a positive, evaluative condition just in case the concept applies only to traits that typically result in actions for which there is a reason. A concept of a character trait has a negative, evaluative condition just in case the concept applies only to traits that typically result in actions that there is reason not to perform.

argue that W1, as portrayed above, is misdescribed. To formulate this response, more must be said about courage. Courage is not exhibited by every act in which the agent overcomes fear. If a soldier endures gunfire in order to retrieve a button that has fallen off her uniform, her act is stupid, not courageous. If I cross a busy highway in order to save ten steps, then my act is stupid, not courageous. These acts involve overcoming fear, but they are not courageous, because an act is courageous only if the goal one hopes to attain is worth the risk involved in performing the action.

What is it for an action to be worth the risk? One popular theory holds that an action is rational if, given the values and probabilities of the outcomes of that action and other actions, the action maximizes the expected utility for the agent. If we accept this theory, we might say that an action is worth the risk if it maximizes expected utility; i.e., an action is worth the risk if it is rational. Thus, on this theory, an

action is courageous only if it is rational.

Totally aside from a detailed theory of rationality, it would be odd to take an action to be worth the risk, but to hold that there is no reason to take the action. So even on a very intuitive account, an action is courageous only if there is a reason for performing it.

If these views are correct, then it seems the concept of courage requires that an act is courageous only if there is a reason for doing it.¹⁰ Then, given the account of evaluative conditions above, the concept of courage has an evaluative condition of application.

Let us reconsider W1. At W1, it would be irrational to take the risk involved in attempting a slave rebellion, since the result will be the life-long torture of all who rebel or benefit from rebellion. So the above description of W1 seems to be off the mark. Some acts, such as rebellion, are worth the risk at other worlds but not worth the risk at W1. Thus, W1 is not a world at which acts of courage lack positive value; rather, it is a world at which acts that would normally be courageous are not courageous because there is no reason for doing them.

We can think of W1 and a world similar to it. At W1, members of the master species are aware of every human action and the psychology behind it. Then it is false that courage exists at W1 and lacks value at W1. Courage does not exist at W1, because it is impossible to

¹⁰ The connection between reason and courage has its origin in Greek thought, and the notion that an action is courageous only if it is worth the risk is, on some interpretations, implicit in Aristotle's account of virtues as means.

act courageously at W1. Courageous actions are impossible at W1, because it is never rational—never worth the risk—to overcome fear at W1. But it may be that, at any world where courage is possible, an agent can perform a courageous action only if there is a reason to perform it.

Now consider this similar world:

World W3: This world is exactly like W1, except that the master species does not detect every human action and the psychology behind it. Specifically, the masters sometimes overlook small movements of the thumbs. Humans develop a set of thumb movements, such that they converse in a "thumb language." The language is greatly valued by humans, since it provides a realm of freedom beyond the control of their masters. The masters occasionally detect free thumb movements, however, in which case they torture the thumb mover and those around her.

Speaking thumb language is courageous at W3. It is true that, at the actual world, thumb movements are rarely courageous, and it is also true that, at W3, attempting rebellion would not be courageous. This shows that the behaviors that constitute courage at W3 differ significantly from the behaviors that constitute courage at the actual world. It does not show that courage lacks positive value at W3, or that recognizing an action as courageous is not, at all worlds, also recognizing it as an action for which there is a reason.

The concept of courage, then, appears to be both world-guided and action-guiding. But a similar response is not available in the case of kindness. Given that we want people to lead fruitful, satisfying lives, we would not value kindness at W2. At W2, the kind person makes a sacrifice disproportionate to the good done. So, at W2, it is best that people do not act kindly. There is no reason for agents at W2 to act kindly.

The concept of courage appears not to be purely descriptive and the concept of kindness appears to be purely descriptive. Where do other virtue and vice concepts fall? Among traditional virtues and vices, we may distinguish between *character* virtues or vices and *moral* virtues or vices. Character virtues, such as courage, temperance, and prudence, are typically exercised when the agent overcomes a specific type of temptation, desire, or emotion which threatens to interfere with the agent's achievement of an end. (In the case of courage, the agent overcomes fear.) These virtues are not primarily directed toward others or toward a principle that directly mentions behavior toward others. This is not to deny that the exercise of these virtues often affects the well-being of others. Likewise,

the exercise of the corresponding vices, such as cowardice, intemperance, and imprudence, can adversely affect others. On the other hand, moral virtues, such as kindness, generosity, and honesty, are primarily directed either toward others or toward principles that directly mention behavior toward others. Their corresponding vices, such as cruelty, greediness, and dishonesty, always have some direct effect on others.

Concepts of character virtues are plausibly construed as action-guiding. Although such virtues are typically exercised when the agent overcomes a temptation, desire, or emotion, the agent must overcome these only to a rational extent. Temperance, for instance, does not require that one never indulge, but only that one not overindulge. That is, it requires that one never indulge to the point at which the negative effects of indulgence fail to be worth the positive gains. Thus, for reasons similar to those discussed in the case of courage, temperance may be held to require that the agent only indulge as much as is rational. So an action is temperate, it may be argued, only if there is a reason for performing it.

Concepts of moral virtues, however, are purely descriptive. Examining these virtues individually would be too lengthy, so I have chosen kindness as a typical example. For any moral virtue, there are possible worlds at which all actions that have the virtue lack positive value. Moreover, the central moral virtues—such as charity, kindness, generosity, and honesty—do not involve a reason-guiding condition of application, since they do not require that an agent perform a certain type of action only to a rational extent. Thus, the antidescriptivist argument that we have been considering applies only to concepts of character virtues.

III.

I claimed that, by spelling out the evaluative condition in terms of action guiding, the antidescriptivist can plausibly argue that concepts of character virtues are evaluative. I now want to argue that this antidescriptivist move will not work.

In the case of courage, consider two kinds of examples. First, suppose that false but justified beliefs cause the agent to be misinformed about the possible outcomes of her action. For instance, suppose that soldiers in an isolated bunker believe, on the basis of recent radio reports, that the war is not over. Suppose that they take risks in their attacks on the enemy, yet (unknown to them) the only possible outcome of their actions is the aggravation of a strong enemy that has already won. The soldiers' actions appear to be worth the risk, but they actually have negative value.

The second kind of case occurs when, due to a mistake in complicated reasoning, an agent is misinformed about the utility of an action. Examples are too complex for detailed description; but we can imagine a general, engaged in sophisticated strategic planning, who places her army in a situation that she believes is worth the risk. She could easily reason incorrectly. Again, the action appears to be worth the risk, even if it is not.

Let us say that an agent acts *reasonably* when (a) she attempts to be rational in deciding to act and (b) she succeeds, for the most part, in actually being rational. An agent can act reasonably even if she is misinformed or makes a bad inference. In our examples, we did not imagine soldiers who fought after receiving messages that the war ended; nor did we imagine the general deciding by the flip of a coin. The imagined agents acted reasonably, yet they were mistaken about the value of their actions.

The concept of courage extends to these cases. The case of the soldiers is the clearest. Soldiers are often misinformed about the position or strength of the enemy, and they are often forced to make quick decisions in confusing situations. Thus, they take actions that they would not view as worth the risk, were they able to reflect carefully on all the relevant facts. Yet soldiers can act courageously in performing such actions.

Thus, the soldiers act courageously even when their actions are not worth the risk. Since an agent can act courageously even if her action is not worth the risk, an agent can act courageously even if she does not act in a way that has any actual chance of promoting a worthwhile end.

The consequences are unfortunate for the antidescriptivist. Consider this world:

World W4: This world is like W1, except that the humans mistakenly believe their world is like W3. That is, they mistakenly believe that some actions (e.g., thumb movements) go undetected by the masters. Further, they have very good grounds for this belief. (The masters allow them to speak thumb language for a few years without being punished. The humans notice that people are missing, but the masters tell them that these people were transferred to other workplaces. The masters force those who have been removed for torture to send letters claiming to be well.)

Humans at W4 will reasonably believe that some courageous actions are worth the risk. They will act courageously, even if courage is not of value, and even if, unbeknownst to those who act courageously, courage leads to torture and is therefore of negative value.

At any world where agents can be reasonable but misinformed about whether actions are worth the risk, courageous actions can lack positive value. Thus, the value of courage depends not only on general social and psychological facts, but also on the extent to which agents are well-informed. At W4, agents are poorly informed. Similarly, if soldiers were usually misinformed about the position and strength of the enemy, but reasonably believed they were well-informed, then they would often perform courageous actions that had unfortunate consequences. In such a world, courage would lack positive value.

Let us reexamine the view that the concept of courage has an action-guiding condition of application. Consider three readings of this condition. On the first, weak reading, the condition requires merely that the agent have some reason, whether or not the reason is a good one. On the second, intermediate reading, the condition requires that the agent have a reason, and that the agent be reasonable in accepting that reason. On the third, strong reading, the condition requires that there be a good reason, where this entails that the action is actually worth the risk.

The case of the soldier who endures gun fire in order to retrieve a button shows that the weak reading is not enough. It would not be enough if she had some reason to retrieve the button—for instance, that it would keep her a bit warmer. Unless the soldier reasonably believed her action was worth the risk, she cannot have been courageous. On the other hand, the cases just gone over show that the strong reading demands too much, because an agent can be courageous even if she is wrong about the risk involved.

The action-guiding component in the concept of courage, then, requires only that there be a reason in the intermediate sense—it requires that the agent be reasonable. But, since it does not require that courage be of generally positive value, there can be worlds at which courage has negative value.

The antidescriptivist could object that what has been shown is precisely that there is a sense in which concepts of character virtues are evaluative. For one may hold that, when evaluation is cashed out in the intermediate sense (that is, in terms of reasonableness), the concept of courage and other concepts of character virtues can be shown to have evaluative conditions of application.

The problem with this is that, from the descriptivist point of view, reasonableness itself need not provide value; so there is no ground for holding that a condition which requires being reasonable is an

evaluative condition. The descriptivist can back up this view by noting that there are possible worlds where being reasonable is generally of great negative value, and where acting on gut intuition, uncontrolled emotion, faith, or in some other unreasonable manner is always of greater value. At W4, being reasonable is of no value whatsoever when contemplating courageous actions. We can, if we wish, imagine that at W4 the masters reward unreasonable behavior, perhaps by rewarding those who act only on emotions. The descriptivist holds that our tendency to believe that reasonableness must have positive value results from our concentration on what we take to be valuable at our world. But reasonableness, like courage and kindness, could turn out to have negative value.

IV.

Two related objections may be made to the strategy adopted so far. The first is that, by spelling out evaluation in terms of prima facie reasons, we can treat virtues as valuable at all worlds. On this view, a concept has a positive, evaluative condition of application just in case it may be applied only to actions for which there are prima facie reasons. Then courage, for instance, is of positive value at W4, because (according to the antidescriptivist's intuitions) there is a prima facie reason for acting courageously at W4. This prima facie reason does not give courage the greatest overall value at W4, because the value of courage is outweighed or overridden by other factors. But courage does have value, even at W4.

The second objection is that courage (for instance) may have intrinsic value at worlds like W4, but that the value is outweighed by the intrinsic value of other factors, such as the negative value of the pain suffered by those tortured. If one takes the first objection to presuppose that an action has intrinsic value if and only if there is a prima facie reason for performing it, then the first objection may be treated as a version of the second. (The second objection, unlike the first, can be made without assuming that evaluation is to be explained in terms of reasons for action.)

The first objection remains unclear until the notion of a prima facie reason is explained. On many readings, this account succumbs to the same problems as accounts in terms of reasonableness. For example, suppose that we say there is a prima facie reason for an action just in case there is available a consideration which the agent,

¹¹ According to some religious conceptions of the relation between faith and reason, we live in this sort of world.

were she aware of it, would take to constitute some reason for the action. Then there could be prima facie reasons for actions of a certain sort, even if actions of that sort always had great negative value.

The central problem with both objections, however, is that they beg the question against descriptivism. The claim that courage and other virtues must be valuable is precisely what is being debated. The descriptivist does not share the intuitions that courageous actions are intrinsically valuable, or that an action is courageous only if there is a prima facie reason for performing it.

Moreover, the descriptivist can offer an explanation of why some people have these antidescriptivist intuitions: they believe themselves to live at a world unlike W4, and so they have developed the belief that courage is always valuable. If they believed that their world was one at which courage was not of greater overall value, they would not have developed the intuitions that there are always prima facie reasons for courageous actions and that courage is intrinsically valuable.

Returning to an earlier example, consider the following world:

World W5: Due to an unknown cause, humans have incredibly strong urges to kick others. Moreover, the shins and calves of humans are extremely sensitive. A kick that would result in a barely perceptible tap at our world would lead to excruciating pain at W5.

At W5, kicking others would have great negative value. Refraining from kicking in situations where one had the strong urge to kick would be an important virtue. Parents, educators, and religious leaders would stress that this virtue must be mastered in order to live a good life. Stories would be told about heros who refrained from kicking. And some philosophers would develop the "intuitions" that the concept of refraining from kicking refers to a trait with intrinsic value, or that the concept must apply only to actions for which there is a prima facie reason. From the descriptivist point of view, they would be mistaken. Clearly, what would make refraining from kicking so valuable at W5 would not be its intrinsic value, but its contribution to the good life of humans, given the total set of facts about the human condition at W5. This contribution to overall value would mislead the philosophers into antidescriptivism regarding the concept of refraining from kicking.

The point here is that, for any descriptive concept, we can imagine

a world at which the property referred to by the concept has great value. So we should not let the fact that we happen to live at a world where certain properties have great value mislead us into holding that these properties have intrinsic value.

I turn below to the "holistic" picture of evaluation on which this argument for descriptivism depends. But I want first to note the consequences of descriptivism for realism about virtues.

V.

There are many senses of realism about a property. I shall note three. First, realism may be the view that the property plays a causal role in the world. Second, realism could be the theory that propositions containing concepts that refer to the property have truth conditions that depend on how the world is. Third, realism could be construed as the view that these truth conditions can transcend our recognitional capacities; so propositions containing concepts that refer to the property can be true or false even if we cannot verify their truth or falsity.

If, as I have argued, concepts of traditional virtues are descriptive, then realism (in any of these senses) can hold true of the properties referred to by virtue concepts, without moral or ethical realism holding true. Suppose, for example, that realism is true of kindness. In other words, suppose that kindness plays a causal role, that propositions containing the concept of kindness have truth conditions, and that these truth conditions can transcend our recognitional capacities. Since the concept of kindness is purely descriptive, it does not follow that there is any ethical or moral property such that either the property plays a causal role in the world, or propositions containing concepts that refer to the property have truth conditions, or such propositions can have truth conditions beyond our recognitional capacity. In short, one can be a realist about properties to be valued (such as kindness) without being a realist about values themselves.

It is worth noting that realism about values does not follow directly from antidescriptivism about virtue concepts. One may argue that, for any virtue concept, the descriptive and evaluative conditions can be separated. One can then hold that realism is true of properties denoted by the descriptive condition, but not true of properties denoted by the evaluative condition. To avoid this line of argument, the moral realist about virtues must argue that, although a virtue concept is both world-guided and action—guiding, it is impossible to sort out independent descriptive and evaluative conditions of application. So a competent agent cannot apply the descriptive part of the

concept to a possible action without taking there to be a reason for the action. 12

If descriptivism is correct, the question of whether descriptive and evaluative conditions can be sorted is beside the point, since there are no evaluative conditions. But the special connection between virtues of character and reasonableness illuminates how one could be mislead into taking virtue concepts to be positively evaluative. If a fully competent agent is considering an action that she takes to be courageous, then she must view the action as reasonable. Generally, an action can be viewed as reasonable even though it is one for which the agent does not give a good reason, where having a good reason entails that the agent is actually correct that there is a reason for the action. Yet, since we are now considering the agent herself as the person who takes the action to be courageous, the agent cannot view the action as one which is reasonable but for which she does not have a good reason. For, if the agent viewed herself as not having a good reason, she could not view herself as reasonable. So, if an agent believes one of her own possible actions is courageous, then she has, from her own point of view, reason to take the action. Thus, the concept of courage is action-guiding, in the sense that agents who believe their own present or possible actions are courageous must also take themselves to have a good reason for so acting. This could mislead someone into holding that the concept has an evaluative condition of application.

We have seen that the concept of courage lacks an evaluative condition, if this means that it must apply to what is valuable. So, outside of the viewpoint that an agent adopts in considering her present or possible actions, she may recognize that the action is not valuable. Thus, she can view her past actions as courageous but of negative value, or she can view the actions of others as courageous but of negative value. (This was how we viewed human actions at W4.)

Concepts of character virtues other than courage can also refer to properties that have negative value. Agents who competently apply such concepts to their present or possible actions cannot at that point in time recognize the negative value of their actions. I am not suggesting that this has actually misled philosophers into holding

 $^{^{12}}$ McDowell argues for this view in "Virtue and Reason" and "Non-cognitivism and Rule-following." $\,$

that these concepts have a positive evaluative condition, but I am suggesting that it could so mislead.¹³

VI.

The argument so far relies on a sort of "holism," according to which whether or not a property is valuable depends on a broad set of social and psychological facts. What we take to be valuable—what we hold there is reason to do, all things considered—depends not only on our conception of particular circumstances, but on our total set of beliefs. Just as we depend on many beliefs in order to predict future experience, so we rely on many beliefs in order to determine what ought to be done. No particular property of an agent or an action can be evaluated independently of a large number of beliefs.

An antidescriptivist might argue against this that to base our account on such extreme examples as W1–W4 is to fail to understand that virtue concepts can be applied only from within a given "point of view," "form of life," "conceptual scheme," or "tradition" in which certain assumptions and claims are basic. Among these assumptions and claims will be general claims about the nature of human beings and their social situation. So we assume that humans are not weakened by acting kindly and that they are not enslaved by a superior species. Such assumptions, it might be said, are "at the base of our language game"; they have the status, as it were, of synthetic a priori truths that guide our moral life. On this view, our concepts apply only to worlds where the assumptions are true. If so, then our concepts are improperly applied to the possible worlds imagined in this paper. When applied properly, they always refer to properties that are valuable.

This kind of move is a matter of much debate. But, even if we allow that this sort of move is sometimes acceptable, there are two reasons for not accepting it in the case of virtue concepts. First, in this case, the appeal to "points of view" (etc.) and associated implicit assumptions is not sufficiently similar to examples where such appeal is more appropriate. Some cases in the literature—for instance, Wittgen-

¹⁴ Wiggins, op. cit., stresses the importance of points of view. Alasdair MacIntyre, in After Virtue (Notre Dame: University Press, 1984), emphasizes the importance of traditions. I do not have space here to examine their arguments in detail.

¹³ I have been arguing, against antidescriptivism, that ordinary virtue concepts are not evaluative; yet philosophers can invent concepts that are both descriptive and evaluative. Nothing said here shows that this is impossible. But, since the concepts are invented, the nonrealist about values may argue that the evaluative and descriptive components can be separated, and then endorse realism only about properties denoted by the descriptive component.

stein's discussion in On Certainty¹⁵ of the special role that claims like "Here is my hand" play in our language game—involve assumptions that are so fundamental that, if we gave them up, we would not know how to make any appropriate move in our language game. We do not feel we could make any legitimate claims if we could not assert the claims in question.

The claims we have considered are not like that. Although it is somewhat difficult, for instance, to imagine that we are enslaved to a superior species, we can imagine transporting our values and our moral point of view into that situation. We can suppose that the master species leaves us alone to develop our own morality and conception of the good life. We would still care about morality, and we would still have our ideals of human flourishing, whatever those amount to in detail. Indeed, it is because we would still have them that we can judge that, even in this situation, acts that would result in the torture of many people would not be valuable. By transporting our point of view into situations that we believe are different from our actual situation, we are able to pass judgment on what we would value in those other situations.

The cases we have considered are thus different from those in which variations between points of view are so radical that, if we were to accept the claims made from another point of view, we would not know how to make any legitimate claims. Many readers of this essay could not, from within their point of view, allow that the seventeenth-century Samurai warrior is ethically correct in killing a peasant who hesitated two seconds before bowing. If we were to accept the judgment "it is morally acceptable for the Samurai to kill the nondeferential peasant," we would, on the Wittgensteinian sort of view under consideration, not know how to continue using our moral language. It is central to our moral language game that human life is worth more than that. If we allowed this judgment, we would no longer know when to say killing is impermissible; we would not know how to talk about rights to life or just and equal treatment. We would, in other words, have to give up much of our moral language game. But the cases considered in this paper cannot be successfully assimilated to the Samurai warrior case.

There is a second, more pragmatic reason for not holding, on the basis of considerations about points of view, that our ordinary con-

¹⁵ G. E. M. Anscombe and G. H. von Wright, trans. (New York: Harper & Row, 1969).

cepts are thick, and for not relying heavily, in our account of the ethical realm, on such thick concepts. ¹⁶ This second reason is based on a recognition that there are variations in moral sensibility between cultures. Thick concepts can only be usefully applied from within a particular point of view; they cannot be interestingly applied in arguments between different points of view. Since, in the actual world, there are important conflicts between points of view, treating ethical concepts as thick tends to cut off rational discussion of value, and it leads to uninteresting, nonexplanatory accounts of value.

Consider revenge, for instance. Many of us do not believe that revenge is of positive value, but in some cultures it is regularly used in positive moral evaluations. Suppose it were argued that the concept of revenge applies only to actions for which there is a reason. Then we would not be able to understand how the concept of revenge is applicable to our world, since we hold that there is usually no reason for performing actions that meet the descriptive condition contained in the concept of revenge. If someone takes it to be necessary that the concept of revenge applies only to acts that are of positive value, then we cannot debate with that person about the value of revengeful acts—the issue is already settled. Either we must invent a new concept which has the descriptive component contained in the concept of revenge but which is drained of its evaluative component, or discussion must simply come to an end because we lack a common descriptive concept that can be used in debate. Treating a concrete ethical concept as thick is simply an effective means for cutting discussion short.

BRUCE W. BROWER

University of Illinois/Chicago

¹⁶ Some philosophers might object that I have given an illegitimate "conceptual analysis." I hold that something like conceptual analysis is acceptable, so long as no claim is made that we have special, a priori access to a realm of meanings. In any case, my argument is in complete accord with the Quinean view that by revising our beliefs, we can apply terms in ways which might have been thought to conflict with "rules of meaning." The pragmatic point made in this section is also consistent with a Quinean view.

RULES FOR A GOOD LANGUAGE*

HY does our language contain words denoting certain classes of things but not others? Why, for instance, are there words for dogs and cats but no words for fast animals or dangerous animals? Or why have words for puppies and kittens but no words for old dogs or cats? Why are there words for hills and mountains but no words for big trees and small trees? Why not have words for wooden furniture, or expensive furniture, or brown furniture? And why not get rid of words like 'chair' and 'bed' and talk in terms of 'furniture one sits on' and 'furniture one sleeps on'? Questions of this sort can obviously be generated without end. I shall call these division questions, for they ask why the words of our language divide up reality one way rather than another.

Our philosophical reaction to division questions is likely to vary from example to example. Let me begin by indicating two important ways

in which these examples may differ.

First, in some examples more than others, our initial intuition is that there is some *serious justification* for why our language divides up reality the way it does. I am assuming a distinction here between justificatory reasons and mere empirical (causal) explanations. Even where the latter appeal to universal or innate human tendencies to divide up reality in certain ways, they need not imply that these tendencies are especially rational or good (the tendencies may derive from the structure of the brain without having anything to do with survival or truth or any other virtue). So my first point is that we are faced with a spectrum of examples varying in the degree to which our initial intuition favors the existence of justificatory reasons.

* I was helped by comments on this paper from George Bealer, Jerry Samet, Fred Sommers, and David Wong. I also benefited greatly from responses by members of the Creighton Club of New York where I presented an early version of the paper.

My work was supported by a Mazer Award from Brandeis University.

Hence, the justificatory issue we are considering here must not be conflated with the famous issue over Whorfian "linguistic relativity," for the latter issue is generally focused on the question whether there are universal or innate division tendencies. The anti-Whorfian (universalist) views of such contemporary psychologists as Eleanor Rosch and Amos Tversky may be compatible with denying any rationale for our division practices. See, e.g., Rosch, "Human Categorization," in Neil Warren, ed., Studies in Cross-cultural Psychology, vol. 1 (London: Academic, 1977), pp. 3–47; and "Linguistic Relativity," in P. N. Johnson-Laird and P. C. Wason, eds., Thinking: Readings in Cognitive Science (New York: Cambridge, 1977), pp. 501–519; Tversky, "Features of Similarity," Psychological Review, LXXXIV, 4 (July 1977): 327–352.

0022-362X/88/8512/0694\$02.40

© 1988 The Journal of Philosophy, Inc.

At the far end of this spectrum are examples of nonstandard ways of dividing up reality which strike us as so utterly strange and bizarre that we can scarcely question their irrationality. These examples are especially worth considering carefully, for they should reveal vividly the general kinds of justificatory considerations which apply to division questions.

I do not think that any example I have so far mentioned belongs to this far end of the strangeness spectrum. A famous example that surely does is 'grue': the division question here is why we should divide up reality in terms of 'green' and 'blue' rather than 'grue' and 'bleen'. Let me mention another example. Imagine a language like English except that, instead of containing the two words 'green' and 'circular', it contains the three words 'gricular', 'grincular', and 'ngricular' which apply, respectively, to things that are either green or circular, things that are either green or not circular, and things that are either not green or circular. For any sentence in ordinary language, there will be a logically equivalent sentence in this language, since 'green' and 'circular' are logically equivalent, respectively, to 'gricular and grincular' and 'gricular and ngricular'.2 I take it that our strong immediate reaction to the Gricular language is to ridicule it as absurd and to suppose that there must surely be a serious rationale for our not operating with such a language. If someone is tempted to suggest that we need not react so strongly to what may only be a highly localized quirk in a language, let me stipulate that, in the Gricular language, numerous ordinary words are replaced in the same way 'green' and 'circular' are. This must intuitively strike anyone as being absurd.3

In what follows, I shall use the Gricular language rather than the Grue language as my model of an ostensibly absurd way of dividing up reality. The fact that 'grue' appears to be "positional"—i.e., it

² 'Logical equivalence' can be understood either in the sense of a priori equivalence or in the sense of metaphysically necessary equivalence. My assumption is that these two senses coincide with respect to every example I shall consider in this paper.

Two assumptions are implicit in this sort of example. First, it must be understood that, in the Gricular language, the occurrence of 'gricular' is not limited to the two expressions 'gricular and grincular' and 'gricular and ngricular'. If that were so, one might even question whether 'gricular' is a word (because it would lack the syntactic and semantic independence characteristic of words), and the distinction between ordinary language and the Gricular language may threaten to collapse. But we should imagine that 'gricular' can show up in any linguistic context accessible to a term denoting things that are either green or circular. Second, 'gricular' is assumed to be an unambiguous word that denotes anything that is either green or circular, rather than being merely ambiguous as between 'green' and 'circular'. Homonyms or ambiguous words are, of course, quite ordinary.

appears to involve an implicit reference to some moment of time—suggests some special problems that distract attention from the more general division questions I want to discuss. It is better for present purposes to consider a language, such as the Gricular language, containing numerous words which, like 'gricular', seem absurdly disjunctive without being positional.

This brings me to the second way in which division questions may differ. The Gricular language divides ordinary things into strange classifications. That is, it contains strange principles of classification. We should also consider examples of languages containing strange principles of individuation whereby reality seems to be

divided up into strange things.

Let me refer to an example I have discussed elsewhere.4 Imagine a language, Contacti, which is like English in many ways except that it bizarrely defines the identity through time of many things in terms of their contact relations. The language will contain, for example, the word 'cperson' (corresponding to the English word 'person') which operates as follows. It is a principle of "cpersonal identity" that during an interval when two "cpersons" come into contact exclusively with each other, they temporarily exchange their physical and mental characteristics. Hence, the history of a single "cperson" will combine what we would ordinarily regard as stages of one person, followed by stages of a second person with whom the first is exclusively in contact, followed by stages of the first person after exclusive contact is broken off. We can imagine Contacti generalized to include many other kinds of bizarre individuals such as "ctables" and "cdogs" whose identities depend in a correlative fashion upon their contact relations.

Contacti sentences are (stipulated to be) logically equivalent to ordinary sentences in the manner indicated. The division question here is why it is better to divide up reality into individuals in accordance with the latter sentences rather than the former.⁵

Our initial intuitions with respect to division questions imply, then, that there are serious rational constraints on how our words should classify and individuate. I shall attempt to formulate what seem to be the most promising constraints in terms of various "rules for a good

⁴ The Concept of Identity (New York: Oxford, 1982), ch. 10. (My presentation of the example here differs in a slight notational respect from that in my book.)

In assessing the various languages, what language should I be using in this paper? Following what I take to be the standard (and perhaps inevitable) practice in such discussions, I adopt a kind of hybrid metalanguage combining English with the strange languages under discussion. Hence, I assert sentences in this paper which could not be asserted in straight English, for example, the sentence, 'Ctables are bizarre things that exchange their qualities when they touch'.

language."⁶ A minimal condition of adequacy for such rules is that they at least exclude the strangest examples. If they succeed in this, then we can try to apply them to more moderate examples. We shall see, however, that it is sufficiently difficult to find convincing rules that even satisfy the minimal condition. This somewhat perplexing negative result will be assessed in the final section of the paper.

It may seem immediately plausible to suppose that the rationality of a classificatory or individuative scheme is, at least to some extent, context dependent, that a scheme may be rational at one time and place but not at another. Whether context might have a bearing even with respect to the strangest examples under consideration is questionable. In any case, I should make it clear that my formulation of the problem begs no questions against a philosopher who wants to emphasize contextual considerations. This philosopher should expect to find rules that prescribe what constitutes a good language relative to various contextual features. Several of the rules I shall be considering readily lend themselves to a contextualist elaboration. The context I have been tacitly assuming, and which I will continue to have foremost in mind, however, is our own. The philosopher who stresses context should expect, at the very least, to be able to explain why the strangest seeming schemes would be irrational or bad in our context. Even this modest expectation, it will turn out, is not easily met. The expectation will not be met if we can imagine people as rational and knowledgeable as us, living in our environment, whose survival and well-being depend, in essentially the same way ours do, on how they physically interact with the environment, and who are doing as well as we are, though they employ the strange schemes.⁷

I have considered elsewhere the question whether the "thoughts" or "meanings" expressed by sentences in ordinary language can be said to be *identical* with those expressed by logically equivalent sentences in the strange languages. 8 Related questions might be raised

⁶ A more accurate title for this paper would be "Rules For How the Words of a Good Language Divide Up Reality," since there are various other kinds of possible "rules for a good language" which I am not considering here, for example, rules related to purely syntactic or formal features of language.

As will emerge, I assume that the expectation cannot be met by trivial (and, in a sense, question-begging) explanations that simply take as given cognitive propensities friendly to the ordinary scheme and unfriendly to the strange schemes (for example, the propensity more readily to notice aspects of the world which answer to the ordinary scheme than those which answer to the strange schemes). No such explanation would tend to show that we cannot imagine people as rational and knowledgeable as us, living in our environment, etc.

⁸ See "Strange Thoughts of the Third Kind," P. H. French, T. E. Uehling, Jr., and H. K. Wettstein, eds., *Midwest Studies in Philosophy*, vol. XII (Minneapolis: Minnesota UP, 1988). pp. 6–8.

with respect to "properties" and "facts." I leave all such questions open here. The burden would be on someone to explain how the answer to these questions might solve our problem. Of course, if the answer is that any thought (meaning, proposition, and so on) expressible in ordinary language is expressible in the strange languages, and vice versa, then it follows that, in *one* sense, the difference between ordinary language and the strange languages is slight. But that makes the problem *worse*. For the problem is to explain our intuition—and it is a basic premise of this discussion that we do have such an intuition—that it would be absurd for us to speak the strange languages.

In treating the strange example, there are two somewhat different questions that might be raised. The first is why ordinary words (such as 'green' and 'circular') could not be replaced by strange words (such as 'gricular', 'grincular', and 'ngricular'). The second question is why the strange words could not be added to our ordinary vocabulary. In what follows, I shall be thinking primarily of the first, more extreme question, though much of what I say could also be applied to the second question. (Neither of these questions should be conflated with W. V. O. Quine's question about the inscrutability of reference. The inscrutability question asks why it is rational to reject strange interpretations of ordinary languages. My question is why it is rational to reject strange languages, interpreting these in as ordinary a way as possible.)

I shall consider five general rules and various subrules. The general rules will appeal, respectively, to (1) metaphysical constraints, (2) epistemological constraints, (3) constraints related to the learning of language, (4) constraints related to explanations, and (5) pragmatic constraints. These distinctions are obviously to some extent artificial and one of my purposes will indeed be to indicate various ways in which the rules are interconnected.

II.

Rule 1 (the metaphysical rule). Language should divide reality at the joints.

Perhaps another formulation of Rule 1 would be: The structure of language should correspond to the structure of the facts.

We want to try to interpret Rule 1 as at least excluding the examples of classificatory and individuative strangeness. For the case of classification, the most promising interpretation, I think, is along the following lines:

1.1 (regarding classification). A general word should denote things that form a similarity class, i.e., a class of (actual and possible) things that

are (sufficiently) more similar to each other (in some respects) than to other things.

To make our assessment of 1.1 a bit more tractable, let us limit our attention to monadic (nonrelational) words. I intend that 'word' in 1.1, as well as in the subsequent discussion, should be understood as "word taken in a particular sense," so that the mere occurrence of homonyms or ambiguous words does not constitute a trivial violation of the rule. The notion of a "similarity class" in 1.1 needs to be further clarified, of course (much will depend on how the parenthetical insertions are worked out), but let me assume that this can be accomplished.⁹

There are at least two crucial questions to raise about 1.1. The first has frequently been discussed in the literature, especially in the writings of Nelson Goodman. This question is whether similarity is an objective relationship independent of language. Is there an objective sense in which circular things form a similarity class but gricular things do not? If the answer to this question is no, then 1.1 cannot get off the ground.¹⁰

The second question is less often discussed, though this is, I think, the more fundamental question. Suppose that there *is* an objective distinction between similarity classes and nonsimilarity classes. How would it then follow that words should denote similarity classes? Since we see in such examples as the Gricular language how reality can be described using words for nonsimilarity classes, what special metaphysical virtue is there in having words denote similarity classes?

Let me temporarily ignore the first question. What can be said in answer to the second? It would be wrong, though tempting, to an-

¹⁰ Perhaps we could still consider the rule: A general word should denote things that form a similarity class relative to the speaker's similarity space. Presumably this could not provide the sort of rationale for our division practices which we are

seeking. (See also my later remarks about salience.)

⁹ Throughout this discussion, I rely on an intuitive notion of a "word (taken in a particular sense)." It is possible that the discussion might be advanced in some non-obvious way by looking more closely at this notion, but I shall not attempt to pursue this line here. (On the notion of a "word," see, for example, Paul Ziff, Semantic Analysis (Ithaca, NY: Cornell, 1960), especially pp. 9–10, 20/1, 76–82; and W. V. Quine, Word and Object (Cambridge: MIT, 1960), pp. 129–131.) A version of 1.1 is stated in David Lewis, "New Work for a Theory of Universals," Australasian Journal of Philosophy, LXI, 4, (December 1983): 344–377, esp. pp. 372–377. (Where 1.1 requires words to denote similarity classes, Lewis requires words to signify "natural properties.") More elaborate versions of 1.1 could be modeled on the detailed theories of such writers as Rosch and Tversky but, as noted above, one should not take for granted that such empirical theories have any normative force.

swer that, if we spoke a language whose words did not denote similarity classes, we could not express our knowledge of similarities (or, more generally, that, if we spoke a language that did not divide reality at the joints, we could not express our knowledge of reality's joints.) Assuming we have such knowledge, we could easily express it in the Gricular language, for example, by saying, "The class of gricular things is not a similarity class, whereas the class of things that are gricular and grincular is a similarity class." The critical question is why the structure of language should have to "show" reality's joints, as Ludwig Wittgenstein required in the *Tractatus* and as 1.1 requires.

If there is any answer to this question, I think it must incorporate at least the most general Tractarian idea that language has certain substantive necessary features which we can know a priori. In behalf of 1.1, it can be claimed that one necessary feature of language is to contain words that serve a classificatory function and, furthermore, that 1.1 gives us what is in some sense the *essence* of classification. To classify just *is* at bottom to compare things, to note resemblances. Hence, at least the most basic (primitive) general words of a language must necessarily place things in similarity classes. 1.1 might be taken to imply that words like 'gricular', which do not denote similarity classes, can be at most derivative or secondary words in a language but they cannot be basic words.

I think that this line of defense of 1.1 is not without force, though I am at present unable to elaborate it further. But let me now go on to raise a third question about 1.1.

How does this rule relate to individuative strangeness? One might be initially inclined to think that, if words must denote similarity classes, then such strange individuative terms as 'ctable' and 'cperson' will automatically be debarred. I think this is a mistake, however. For the class of cpersons (ctables) is a similarity class; cpersons are as similar to each other as persons are similar to each other. I am not making the trivial point that cpersons viewed at a moment are indistinguishable from persons viewed at a moment. The relevant question is whether (the history of) cpersons viewed as four dimensional space-time entities constitute a similarity class. I am suggesting that they do (if persons do). Cpersons are entities that are otherwise like persons except for the bizarre transformations when they come into contact. Certainly such entities seem to form a strikingly similar class. So even if we accept 1.1 we still appear to have no argument at all against individuative strangeness.

The connection between similarity and individuation is complicated by the fact that in some literature salience is taken to be an ingredient of similarity.¹¹ It might be suggested that cpersons do not form a similarity class because cpersons are not (sufficiently) salient. The appeal would then be to 1.1 understood as implying:

1.11. A general word should denote things that form a salient similarity class.

It is far from clear what exactly is meant by saying that cpersons are not salient (since contact perhaps is salient¹²), but I assume that there is some important sense in which this is correct. Nevertheless, it is futile to appeal to 1.11 as an argument in behalf of our ordinary practices for dividing up reality. For salience is clearly *subjective*, and there is nothing in 1.11 to suggest why it would be *irrational* or bad for such strange items as cpersons to appear salient to us (i.e., to be readily noticed by us). If there are pragmatic or other reasons why this would be irrational or bad, these reasons need to be brought out. 1.11 provides no such reasons.

Hence, if our only appeal is to 1.11, we have in effect given up hope of finding a rationale for our division practices. In our original appeal to 1.1, the hope was to be able to regard similarity as objective and to claim that there is a constraint on language requiring words to be aligned to the objective similarities. But this presupposes that similarity is *not* dependent on salience. If there can be objective similarity judgments, they must somehow take into account every relevant feature of a situation, whether salient or not. Only on these terms could the appeal to 1.1 possibly succeed in providing a defense of our classificatory practices. It follows that 1.1 cannot possibly succeed in defending our individuative practices, since it is clear on reflection that (if similarity is objective) the world must contain untold many strange things, such as cpersons, that form (nonsalient) similarity classes.

To deal with the individuative case, we apparently need a rule that stands to individuation in the way that 1.1 stands to classification. I think we find, however, that no such rule is readily forthcoming. The rule could not simply say, "A general words should denote *ordinary things*." This would merely beg the question. What we are looking for is some explanation of why our ordinary principles of individuation are metaphysically special, in the way that 1.1 (assuming similar-

¹¹ See, e.g., Tversky, "Features of Similarity," pp. 340-343, and Quine, *The Roots of Reference* (Chicago: Open Court, 1973), p. 25.

¹² Evidence that contact is salient at least in infancy is found in Elazebeth Spelke, "Perception of Unity, Persistence, and Identity: Thoughts on Infant's Conception of Objects," in J. Mehler, ed., *Neonate Cognition* (Hillsdale, NJ: Erlbaum, 1985), pp. 95–97.

ity as objective) appears to offer some explanation of why our ordinary principles of classification are metaphysically special. Now, in many current accounts,13 our ordinary principles of individuation for bodies are analyzed in terms of a highly complex set of conditions. There are, first, various conditions of connectedness and boundedness which define a thing's unity at a given moment, then there are a number of other conditions of continuity which define a thing's identity through time, and all of these conditions are somehow linked up to certain so-called "sortal" terms satisfying some conditions (the "conditions" that define the "sortals" may indeed consist in little more than a list of terms which so qualify). Even the special case of personal identity has been defined in terms of complex connections and continuities in much recent literature (e.g., in Derek Parfit and Robert Nozick). If accounts of this general nature are correct, the rule we are looking for might take the following schematic form:

1.2 (regarding individuation). A general word should denote things whose contemporaneous parts stand in the complex interrelations X and whose successive stages stand in the complex interrelations Y (with respect to "sortal" terms satisfying conditions Z).

One immediate problem for 1.2 is whether the schematized conditions of individuation could possibly encompass the variety of "nonconcrete" things (such as events, properties, conditions) denoted by many ordinary general words. Even if one attends exclusively to "concrete" things, when one seriously studies the complexity of conditions which seem to constitute our ordinary individuation of such things, it becomes difficult to sustain the judgment that our individuative practice has any metaphysical distinction. It is difficult to believe that the particular concoction of conditions required by 1.2 defines a sense of individuation which is essential to any language, in the way that, in 1.1, the similarity class *can* perhaps define a sense of classification which is essential to any language. If 1.1 provides us with a vaguely defensible metaphysical basis for criticizing strange classifications, 1.2 seems to do worse with respect to individuative strangeness.¹⁴

¹³ Such as the one I offer in *The Concept of Identity*, op. cit., chs. 1–4. ¹⁴ I here ignore another kind of argument against individuative strangeness, one that derives from the ontological premise that, while there do exist such ordinary things as persons and tables, there do not exist such strange things as cpersons and ctables. I think that this kind of argument begs the question in an unilluminating way, for it evidently could not be expressed in Contacti (nor in the hybrid language I employ in this paper). For further discussion of this point, see "Strange Thoughts of the Third Kind," pp. 12–14.

III

Rule 1 appears not to have been decisively helpful. Let us pass on to the other rules, some of which can be covered more quickly. The other rules need not be kept sharply separate from Rule 1. Indeed, if one believes in "reality's joints," the other rules can be construed as explaining what good is served by the practice of dividing reality at the joints (whereas Rule 1 was taken to imply that this practice is somehow *intrinsically* good or necessary, not for some other reason.)

Rule 2 (the epistemological rule). Language should allow for the formulation of correct inferences.

The idea will now be that the strange languages would run afoul of this epistemological rule, that speaking such a language would lead to our making incorrect inferences.

The obvious question that must be raised is why speaking a strange language should significantly affect our inferences. Presumably the sort of inference which might be thought to be affected is inductive inference. But suppose that we formulate in ordinary language a correct inductive inference from the composite premise O_1 to the conclusion O_2 . In the strange languages under consideration there will be sentences S_1 and S_2 logically equivalent, respectively, to O_1 and O_2 . Given a very plausible "equivalence principle" for inductive inferences, it follows that the "strange inference" from S_1 to S_2 would be just a correct as the ordinary inference from O_1 to O_2 . How, then, could speaking a strange language affect our inferences?

Someone who believes that the strange languages violate Rule 2 is likely to have in mind a subrule such as the following:

2.1. A general word should be projectible.

A word is projectible, in the sense here intended, if observations with respect to any (sufficiently large) subset of its extension provides inductive evidence about the rest of the extension. I assume that data collected with respect to any (sufficiently large) subset of green things, even if all of these be dark green or soft and green, provide us with *some* selective evidence for a generalization about *all* green things, even the light green things or the hard green things (that is, a generalization about the light green things or the hard green things will be more confirmed than a generalization about some arbitrary set of things.) Hence, 'green' is projectible. On the other hand, data collected with respect to the green gricular things will presumably Provide no selective evidence whatever for a generalization about the circular gricular things (that is, a generalization about the circular

things will be no more confirmed than a generalization about any

arbitrary set of things.) So 'gricular' is not projectible.

If we accept 2.1, together with our ordinary judgments about projectibility, strange words like 'gricular' are ruled out. I shall not question our ordinary judgments about projectibility, but what is the argument for accepting 2.1? Why would it be an epistemological problem if expressions rather than words are projectible in a language? Imagine a group of Gricular speakers who project not the word 'gricular' but the expression 'gricular and grincular'. That is what they ought to do (given our ordinary judgments about projectibility), so why would they not? Just as we project 'green' but not 'green or circular', they will project 'gricular and grincular' (equivalent to 'green') but not 'gricular' (equivalent to 'green or circular'). Why should this be an epistemological problem for them? What sort of "bewitchment of intelligence by language" is supposed to get them to project "gricular" irrationally?

According to Goodman's theory of entrenchment, a term is projectible to the extent that it is well entrenched. Rule 2.1 would then imply that the words of a language must be well entrenched. Why should that be? What epistemological harm would it do if none of the words were entrenched but only certain expressions were?¹⁵

I do think that there are some views about projectibility from which 2.1 might follow. It might be held that a word is projectible if and only if it denotes a similarity class. (This is close to Quine's view in "Natural Kinds."16) If we accept this view and also accept Rule 1.1, which requires general words to denote similarity classes, then it will indeed follow that general words must be projectible. But, if 2.1 is merely derived from 1.1, it does not provide us with an independent argument against strangeness. We have no epistemological considerations as such which count against the strange languages.

I have one more point to make about 2.1. If I was right in suggesting earlier that (assuming similarity as objective) strange individuative terms like 'ctable' and 'cperson' denote (nonsalient) similarity classes, there is apparently nothing to prevent these words from being projectible. Then, even if one accepts 2.1, there is still no argument against individuative strangeness.

An epistemological argument against individuative strangeness might be thought to derive from some such rule as the following:

125 ff.

¹⁵ Compare with Nelson Goodman, Ways of Worldmaking (Indianapolis: Hackett, 1978), pp. 128/9. Goodman is evidently addressing the same sort of question I have just raised but it is not clear to me what position he is taking on it. ¹⁶ In Ontological Relativity and Other Essays (New York: Columbia, 1969), pp.

2.2 (regarding individuation). Principles of individuation should be such as to generate correct inferences having the linguistic form, "This thing of sort S has been F in the past. Therefore, it will (probably) be G in the future" (where 'F' may or may not be the same as 'G').

It seems that ordinary things often satisfy 2.2 (at least where 'F' and 'G' represent "intrinsic" properties.) Strange things like ctables and cpersons perhaps do not. You might not even be able to predict, for instance, whether a given male cperson will remain male for the next few moments. (But how clear is the contrast to the case, for example, where you cannot predict what shape a given piece of clay will have for the next few moments?)

Apart from difficulties of detail in applying 2.2 to cases, the really important question to ask is why we should accept any rule like 2.2. What is the epistemological virtue of the linguistic form cited in 2.2? If we spoke a strange language we would have a logically equivalent inference, but one bearing a different linguistic form. Why should that be worse?¹⁷

IV.

It seems unclear, in sum, how the epistemological consideration expressed by Rule 2 can serve to justify our ordinary ways of dividing up reality. The issue of projectibility, however, connects up to another consideration that may seem important to division questions, namely, the demand that the words of a language be *learnable*. It may seem correct to suggest that the strange languages would not be learnable, or at least not in the way that ordinary languages are learnable.

Let us then have as our next general rule:

Rule 3 (the language learning rule). Language should be learnable.

Of course, in one sense the strange languages evidently are learnable, since anyone reading this paper has learned a few of them. The relevant question, presumably, is whether such words as 'gricular' and 'ctable' could be learned ostensively in the way that 'green' and 'table' normally are. Now, the notion of "ostensive learning" is problematical at least in its details; one may surely question whether there is any neat division between words that are "learned ostensively" and those which are learned "by definition." But let us assume that a working distinction of this sort can be sustained and see how an argument might be developed on that basis.

¹⁷ Just as 2.1 follows from 1.1 together with certain assumption, 2.2 might follow from 1.2 together with certain assumptions. Again, this gives us no independent epistemological argument.

It may be claimed that to learn a word ostensively is to project a hypothesis about the word's application. But to do so requires that one should be prepared to treat the word as projectible. Hence, strange words that cannot reasonably be treated as projectible could not be learned ostensively.

The appeal is to the following principle:

- 3.1. Only projectible words are ostensively learnable. 18
- 3.1 is false, however. It is false as an a priori normative principle, which is the crucial point for the present discussion. But it is even false as an empirical principle about human learning. Any reasonably intelligent person can be taught a word like 'gricular' ostensively. There is a familiar psychological experiment in which a subject is shown a series of cards containing figures of varying shapes and colors. If the subject observes that the word 'gricular' is applied to some of these figures but not to others, he will eventually catch on that 'gricular' applies to the figures that are green or circular. This is ostensive learning. 19

The spurious appeal of 3.1 derives, I think, from conflating it with another principle:

- 3.11. If the word 'F' has extension E and someone ostensively learns 'F' by being presented with the subset O of E, then the hypothesis that 'F' applies to all members of E must be (regarded by that person as) confirmed on the evidence that 'F' applies to all members of O.
- 3.11 implies that someone can ostensively learn 'gricular' only if he is presented with both green and circular specimens of "gricular." If, instead, he is presented with only green specimens, he would have no evidence for the hypothesis that 'gricular' applies to circular things. 3.11 does not entail 3.1 and there is evidently no argument in 3.11 against the ostensive learnability of 'gricular'.²⁰

¹⁸ Since, as I have claimed, such words as 'ctable' may be projectible, 3.1 may not present an argument against individuative strangeness. One could attempt to formulate a principle that stands to 3.1 in the way that 2.2 stands to 2.1. This principle would fare no better than 3.1.

¹⁹ See Jerry Fodor, *The Language of Thought* (New York: Crowell, 1975), pp. 34–42. Fodor does mention some evidence (pp. 40/1) which may indicate that it is more difficult for human beings to learn words that are equivalent to disjunctions of ordinary words, than words that are equivalent to conjunctions of ordinary words. If this is merely an empirical fact about the human brain, it has no clear bearing on the present discussion.

Here is one point where the "positionality" of 'grue' makes a difference. It would apparently follow from 3.11 that someone could not ostensively learn 'grue' prior to the year 2000 (assuming this as the time with respect to which 'grue' is defined) without in effect treating 'grue' as projectible. It should be noted that even 3.11 is far from being incontestable. This rule assumes an inferential model of ostensive learning, according to which the learner is viewed as in effect making

The notion of ostensive learnability is connected in the empiricist literature with another somewhat shaky notion: observability. Empiricists have implied that in order for a word to be learned ostensively it must represent things and properties that are observed. Given this assumption, the fact that an intelligent adult can ostensively learn 'gricular' shows that he observes whether something is gricular. Nevertheless, it might be urged that, in the normal course of events outside of experimental setups, people do not observe whether things are gricular, and this is why 'gricular' could not enter a child's language in the normal way. I think it is fairly clear that this cannot provide us with the kind of argument we are seeking, but, to bring the question out into the open, let me express the following principle:

3.2. In order for a word 'F' to be ostensively learned in the normal course of language learning, people must normally observe that things are F.

Let us assume that 3.2 is correct, and also that, in some important sense, we do not normally observe the strange properties and things. The relevant question, however, is why it would be irrational if we did. Let me emphasize that I am not demanding an argument to show why it would be irrational for a creature with some alien sensory apparatus to divide up reality differently from the way we do. It would be a sufficiently interesting result, if it could be shown that our ordinary division practices are at least essential for any creature with our basic sensory apparatus, i.e., a creature who has sight and touch, whose visual field has the same size and dimensions as ours, and so on (a creature with the same "sensations" as ours). But even the limited argument seems hard to come by. Why could there not be a creature with essentially our sensory apparatus who normally observes that things are gricular, or normally observes that two ctables exchange their qualities, and hence readily learns 'gricular' and 'ctable'? The correct answer to this question, at any rate, is not given by 3.2.21

It may be felt that special problems arise with respect to the learn-ability of the first-person pronoun, problems which militate against a strange concept like "cpersonal identity." There are probably several important intuitions at stake here, but I think the most essential one

inductive generalizations about the applicability of a word. This model is highly problematical, as Fodor has argued in *The Language of Thought*, op. cit., pp. 59-64

²¹ The notion of "what is normally observed" is obviously closely related to "salience," and both notions are unhelpful to the present discussion for the same reasons.

has to do with a connection between one's understanding of the word 'I' and memory. Suppose that I am alone in my room now recalling a moment yesterday on the train when I experienced a pain in my tooth. I cannot remember whether at that moment I may have been (exclusively) in contact with someone. If I spoke Contacti (in which an utterance of 'I' refers to the cperson who utters it) I would have to say: "I remember that toothache but I can't remember whether the cperson who had it was me." One may feel that there is a fundamental problem here, that a language that functions in that way could not possibly be learned as a natural language.

Let us consider the following principle:

3.3 (regarding self-identity). In order for the first-person pronoun to be learned in the normal course of language learning, principles of self-identity must be such as to make it normally correct to say something having the linguistic form, "I remember my own past experiences (and not someone else's past experiences)."

Is 3.3 plausible as an a priori or normative principle? (As always, a merely empirical principle will not give us what we want.) I think it is difficult on reflection to see why it is. What is required is an argument to show why the ordinary linguistic form cited in 3.3 enjoys some special distinction. It would seem merely superficial to defend 3.3 on the grounds that the meaning of 'remember' dictates that 'I remember the experience E' entails 'I experienced E'. If this entailment holds in English, we can stipulate that in Contacti it does not hold. Or we can replace 'remember' in Contacti with 'quasi-remember' as this notion has been employed in the recent literature.²² Then, if I spoke Contacti I would say, "I quasi-remember that toothache, but it may not have been mine." Why would it be in principle impossible (or irrational) for a child to acquire the use of 'I' in the context of such Contacti statements?

V.

The notion of "reality's joints" is frequently connected in the literature to the ideal of explaining reality. This suggests that our explanatory ideals may provide a good reason for dividing reality as we do. The appeal is to:

Rule 4 (the explanation rule). Language should allow for the formulation of correct explanations.

²² Quasi memory contains all that there is in memory minus the purported entailment. See Sydney Shoemaker, "Persons and Their Pasts," *American Philosophical Quarterly*, VII (1970): 269–285.

As a defense of our ordinary division practices, Rule 4 seems open to the same general objection that applied to Rule 2. There I appealed to an equivalence principle for inferences. There seems also to be an equivalence principle for (empirical) explanations. If the truth of O_1 explains the truth of O_2 , where these are ordinary empirical sentences, then it seems that the truth of S_1 will explain the truth of S_2 , where S_1 and S_2 are logically equivalent, respectively, to O_1 and O_2 . If this is correct, the strange languages would allow for "strange explanations" that might be as good as their ordinary equivalents.

Unless it can be shown why the strange explanations are inadequate, we cannot (on explanatory grounds) subscribe to such principles as the following:

- 4.1. A general word should figure as subject or predicate in many laws having the linguistic form, "All (or most) F are G."
- 4.2 (regarding individuation). Principles of individuation should be such as to generate many laws having the linguistic form, "Things of sort S (generally) change and develop in such-and-such a way."²⁴

Perhaps ordinary language satisfies such principles and the strange languages do not.²⁵ But why should this count against the latter? What is the special explanatory virtue of the ordinary linguistic forms cited in these principles? Suppose, for instance, that we have an explanation in terms of the rough law "Ripe tomatoes are green." In the Gricular language we would then apparently have an

²³ The principle of explanatory equivalence seems to be implicitly assumed in virtually all of the literature. An exception is Peter Achinstein, *The Nature of Explanation* (New York: Oxford, 1983), pp. 363–364, but Achinstein rejects the principle only for certain special cases that appear to have no bearing on the present discussion. Of course, there is no equivalence principle for "mathematical explanation," but that seems to be a quite different story. It should be noted that the explanatory equivalence principle does not appear to conflict with Hilary Putnam's requirement that superfluous information must not be added on to adequate explanations, since replacing an explanation by a logically equivalent one does not add any information; see his "Reductionism and the Nature of Psychology," in J. Haugeland, ed, *Mind Design* (Montgomery, VT: Bradford, 1981), pp. 206/7.

²⁴ 4.1 seems to be presupposed with varying degrees of explicitness in any number of discussions in both the philosophical and psychological literature (see, e.g., the papers by Rosch and Tversky cited above). 4.2 is strongly suggested in David Wiggins, Sameness and Substance (Cambridge: Harvard, 1980), esp. pp. 70, 188

²⁵ Actually it is problematical how 4.1 would fail to be satisfied by the strange languages. For example, if there are laws of the form "All (or most) green things are F" and "All (or most) circular things are G," will there not be a law of the form "All (or most) gricular things are F or G"? But that is not the main question to stress here.

equivalent explanation in terms of the law "Ripe tomatoes are gricular and grincular." How can it matter to our explanations which

linguistic forms we employ?

This question is not about the "pragmatics of explanation" in the sense of the cultural and psychological conditions that induce people to "feel enlightened" by one explanation rather than another. Obviously we cannot feel enlightened by explanations couched in the strange languages. But we are trying to imagine people for whom these languages are natural and who do feel enlightened by such explanations. Our question is whether they would be wrong to feel enlightened, even though their explanations are logically equivalent to ours. It is difficult to see how this could be.

It might be suggested that the principle of explanatory equivalence should be given up in favor of a certain relevance requirement.26 The idea, put very roughly, is that an explanation like "The tomato is gricular and grincular because it is ripe," by bringing in the irrelevant property of circularity, wrongly implies that there is some connection between the tomato's ripeness and its circularity. I cannot attempt to work out this suggestion in any detail here, but let me indicate one reason why I am not optimistic about its prospects. In the Gricular language, there is apparently nothing to prevent us from explicitly spelling out what is relevant to what. This should take care of any "relevance requirement." For example, we could say, "The tomato is gricular and grincular because it is ripe, but its being ripe is irrelevant to its being ngricular in addition to gricular." This is equivalent to saying that the tomato is green because it is ripe, but its being ripe is irrelevant to its being circular. What more could one want?

In the strange languages we could formulate explanations logically equivalent to ordinary ones; we could explicitly state which states of affairs are relevant to which; and (to recall an earlier point) we could describe "reality's joints," if such a notion makes sense. It is difficult to see what could be lacking in the way of explanations. I think that, really, only one thing could be lacking: The linguistic structures of the explanations fail to "show" (to "picture") reality in the right sort of way. But this point would seem only to lead us back to the discussion of Rule 1.

Clearly, more would need to be said about the explanatory requirement than the rather superficial comments I have here expressed. Perhaps my question about the explanatory equivalence principle may prompt others to clarify the status of this principle.

²⁶ A suggestion of this sort was made by Thomas McKay.

VI.

The final rule to be considered is:

 $Rule\ 5$ (the pragmatic rule). Language should serve the practical needs of its speakers.

We are now to be thinking of "practical needs" unrelated to inductive generalization, or ostensive learning, or any of the considerations that pertain to previous rules. The immediate question that needs to be asked is why the strange languages could not serve our practical needs. Since we could utter sentences in those languages logically equivalent to the sentences we ordinarily utter, why would we not be able to lead our normal lives while speaking such languages?

Perhaps our first temptation is to appeal to the following rule:

5.1. Language should allow for brevity (and simplicity) of thought and communication.

It is not clear to me how a condition of "simplicity" can be genuinely distinguished in the present context from considerations of brevity. In any case, let me concentrate on brevity. The point would then be that, if we tried to express ourselves in the strange languages, we would require long sentences, thereby violating 5.1

One may question the significance of brevity as a general constraint on language. Given the preponderance of chitchat and small talk in human discourse, it is doubtful whether brevity as such is typically a virtue.²⁷ Even in those cases where it is a virtue, it seems questionable that it is sufficiently serious and important to supply the sort of argument we are seeking. Perhaps we can say: Brevity is important, sometimes, to some degree. Could that be our rational basis for dividing up reality as we do?

Let me frame this question in a somewhat different perspective. Imagine that we can assign to any language a brevity index, which is higher to the degree that the language allows for more brevity. Evidently the brevity index of any language can always be improved upon by adding abbreviatory vocabulary (thereby yielding, in the

²⁷ Contrary to what seems to be implied in H. P. Grice, "Logic and Conversation," in Donald Davidson and Gilbert Harman, eds., *The Logic of Grammar* (Encino, CA: Dickinson, 1975), p. 67. Grice has as one of his "conversational maxims": "Be brief (avoid unnecessary prolixity)." Imagine that you come home in the evening and your spouse says, "How was your day?" I take the case of telling someone how your day was to constitute an essential paradigm of the "informative use" of language. Obviously in such a case there is no presumption to "be brief." ("How was your day? Be brief!") With respect to such cases, one could just as well have, in place of Grice's maxim, "Do not be brief (avoid unnecessary curtness.")

relevant sense, a different language.) Now imagine that all possible languages (that could describe the sort of world we experience) are arranged in a series of ascending brevity indexes. English will occupy an undistinguished place in this series, as will the strange languages we have considered, which will be somewhere below. There will be untold many languages above English and untold many languages below the strange languages. How could it be rational to choose one undistinguished point in the series but irrational to choose another one?

It might be answered that there is in any language a trade off between sentential brevity and lexical economy. The latter is measured simply by the number of words in a language. The former is measured more vaguely in terms of the lengths of the sentences that are normally uttered in the language. It may be admitted that we can of course always improve the brevity index of a language by adding vocabulary and thereby diminishing the economy index. And perhaps we cannot say in general what the ideal trade off is between sentential brevity and lexical economy. Nevertheless, we can at least say the following:

5.11. If L has a higher brevity index than L' and L' does not have a higher economy index than L, then L is pragmatically better than L'.

Thus, it would be irrational to choose the strange languages, because their brevity indexes are lower than those of ordinary languages without their economy indexes being higher. (In some cases, such as the Gricular language, even the economy index is lower, because the two words 'green' and 'circular' are replaced by the three words 'gricular', 'grincular', and 'ngricular'.

My doubts about the general significance of brevity makes me have reservations even about 5.11 (imagine someone who says that one board game is better than another whenever the first takes less time to play unless the second has fewer pieces), but perhaps this rule does provide some kind of critique of the strange languages I have been considering. One might suppose that 5.11 will militate against any strange language roughly akin to these. This turns out, surprisingly, to be false. Many strange languages similar to the Gricular language would in fact constitute a great gain in lexical economy.

Imagine a language in which the seven color words 'green', 'red', 'yellow', 'blue', 'orange', 'white', and 'purple' are replaced by three words 'A', 'B', and 'C' as follows. 'A' denotes anything that is green or red or blue or orange, 'B' denotes anything that is red or yellow or orange or white, and 'C' denotes anything that is blue or orange or white or purple. Then 'green' is logically equivalent to 'A and non-B

and non-C', 'red' to 'A and B and non-C', 'yellow' to 'non-A and B and non-C', 'blue' to 'A and non-B and C', 'orange' to 'A and B and C', 'white' to 'non-A and B and C', and 'purple' to 'non-A and non-B and C'.

In general, any $2^n - 1$ mutually incompatible words can be replaced by n words in the manner of the above example. ²⁸ If lexical economy is what we wanted we would choose our words precisely not to "divide reality at the joints." Our words would be equivalent to complex disjunctions of our present words. ²⁹ Hence, 5.11 leaves it open that the crazy color scheme described above might be "pragmatically better" than ordinary English. It must be understood that I am not making an empirical claim about how human children would take to the crazy scheme; I assume they would take very badly to it. ³⁰ But, as far as our present discussion has been able to reveal, we cannot explain why it is good for humans to (be innately disposed to) reject the crazy scheme by saying that that scheme would be "impractical."

The application of 5.11 to other examples is even more questionable. Suppose the three words 'car', 'apple', and 'ocean' are replaced by the two words 'capple' and 'apcean' which denote, respectively, anything that is either a car or an apple and anything that is either an apple or an ocean. Then 'capple and nonapcean', 'capple and apcean', and 'noncapple and apcean' would be logically equivalent, respectively, to 'car', 'apple', and 'ocean'. This language would not only have a smaller vocabulary, in most ordinary instances it would not even require longer sentences, for the context would generally determine if the thing being called a capple is also an apcean, and if the thing being called an apcean is also a capple. It certainly seems arguable that a rational trade off between sentential brevity and lexical economy might *favor* this language.³¹

²⁸ A proof of this is given in my "Negativity and Complexity: Some Logical

Considerations" (in manuscript).

29 One could achieve a form of lexical economy by employing the device described in Quine, "Reduction to a Dyadic Predicate," The Journal of Symbolic Logic, XIX, 3 (September 1954): 180–182. However, Quine's device requires the replacement of monadic words denoting particular things by a relational word denoting highly complex set-theoretical items. Perhaps we can take it for granted that Quine's device could not possibly enter language at the same level as ordinary words.

³⁰ See Rosch, "Human Categorization," op. cit., pp. 3-15.

³¹ Despite frequent allusions in the psychological literature (e.g., in the cited papers by Rosch and Tversky) to the "(cognitive) economy," or "efficiency," or "(optimal) informativeness" of certain classificatory schemes, I cannot find any substantive account of how the capacities of the human brain might encourage one particular kind of trade off between lexical economy and sentential brevity. Even if

It may, however, be felt that considerations of brevity and economy do not really capture what seems profoundly "impractical" about the strange languages. Is there not a sense in which those languages would fail to reflect what is important and interesting to us, whereas ordinary language does reflect this? One may want to appeal to such rules as the following:

- 5.2 (regarding classification). A general word 'F' should be such that it is often important to us whether something is F.
- 5.3 (regarding individuation). A general word should denote things whose changes are often important to us.32

I have two questions to raise about this. First, I think it is far from obvious what can be meant by saying that ordinary properties and changes are important to us (or interest us) in some way that the strange properties and changes do not. The fact that something is green, taken without any collateral information, will generally be of no importance to anyone. On the other hand, the fact that something is gricular, taken together with some relevant collateral information (including, e.g., the information that the thing is not circular), will often be of importance. How can one say, then, that green but not gricular is important to us? (The following may illustrate the difficulty. Given that a particular object is a tie I need to wear with a certain outfit, it might be important to me that the object is green. Since a tie is green if and only if it is gricular, it may seem equally correct to say that, relative to the given information, it is important to me that the object is gricular.)

This was a question about the applicability of 5.2 and 5.3 to cases. But the main question I want to emphasize is why we should accept these rules. Suppose that green is important and gricular is not. How does that militate against the Gricular language? What is the virtue of representing important properties and changes with single words rather than complex expressions? If someone says "It's gricular and

(Cambridge: Harvard, 1981), pp. 46, 84-86.

See, e.g., Variations of 5.2 and 5.3 are prevalent in pragmatist literature. See, e.g., William James, The Principles of Psychology (New York: Holt, 1890), I, pp. 284-290; C. I. Lewis, Mind and the World Order (New York: Dover, 1929), pp. 49-52; Stuart Hampshire, Thought and Action (London: Chatto & Windus, 1959),

such an account were forthcoming, a pragmatic argument for our division practices could not fully succeed, I think, unless it could be shown why it is good for creatures who live the sorts of lives we do to have brain capacities that encourage that particular tradeoff. See also Robert Nozick's perplexing claims in behalf of the "informativeness" of the closest continuer schema, in Philosophical Explanations

grincular" to express an important fact, why would that be impractical (unless one goes back again to brevity considerations)?³³

The second question applies even to the case of self-identity. It does seem evident that changes in a person are especially important to that person, whereas, if we spoke Contacti (and continued to live the way we do), we could not say that changes in a cperson are especially important to that cperson. In a number of recent publications there seems to be suggested a rule like the following:

5.4 (regarding self-identity). Principles of self-identity should be such as to make it correct to say something having the linguistic form, "I care primarily about my own future (and not someone else's future.)"³⁴

But the question that needs always to be squarely faced is what the special *virtue* can be of some particular linguistic form—in this case, the linguistic form cited in 5.4. If one spoke Contacti, one would replace that linguistic form by a logically equivalent one. How could that be impractical (unless, again, we revert to brevity considerations)?

The impulse to appeal to pragmatist considerations in this discussion is extremely strong and some readers will remain unconvinced by my skeptical questions about such considerations. Perhaps it is worth emphasizing once again that my questions are directed toward normative principles rather than empirical laws. There may well be empirical laws describing some natural alignment between the sort of language people employ and the sort of needs they have in a given socio-historical context. Indeed, one might attempt to recast rules 5.2, 5.3, and 5.4 as approximations to such empirical laws. (Some previous rules might possibly be recast as describing the natural

³⁴ See Sydney Shoemaker, "Comments," in N. S. Care and R. H. Grimm, eds., Perception and Personal Identity (Cleveland: Case Western Reserve, 1969), pp. 123/4; Nozick, Philosophical Explanations, op. cit., pp. 108-110; Derek Parfit,

Reasons and Persons (New York: Oxford, 1984), pp. 292/3.

³³ I am here assuming that, if the fact that the thing is green is important, so is the logically equivalent fact that it is gricular and grincular (leaving it open whether these facts are to be regarded as identical). It should be clear that, even if one supposes that the important fact is that things are gricular and grincular, rather than that they are gricular, it would not follow that 'gricular' would occur only within the expressions 'gricular and grincular' or 'gricular and ngricular' (thereby perhaps threatening the status of 'gricular' as a word; see fn. 3.) Speakers of the Gricular language would regard 'It's gricular' as a complete sentence capable of expressing a truth—albeit perhaps an unimportant one—and there will be correlative general truths, such as, "Peppers and tomatoes are gricular." Even in practical terms, in many typical examples, such as that of the tie, it would suffice to say "It's gricular," the context supplying the rest of the needed information. This point is even clearer, perhaps, for the example of 'capple' and 'apcean' since, as noted above, when one of these words is used, the context would almost always make it unnecessary also to use the other word.

alignment between people's language and various aspects of their beliefs and theories.) I am not addressing the question whether there are such empirically natural alignments. My question concerns the rules only when they are interpreted as having normative force, as prescribing a rational alignment between one's language and one's (contextually determined) needs. The pragmatist must take care not to trivialize the issue by invoking question-begging "needs" such as, perhaps, the second-order need to have the natural alignment between one's language and needs. The relevant question, surely, is what good is served by the natural alignment. We should compare ourselves to imaginary people who do not have the propensity to align their language to their needs in the way we do. How could the fact that we do have this propensity help us, in comparison to them, to eat better, to survive longer, to procreate better, to enjoy ourselves more, and so on? I am not acquainted with any pragmatist literature that deals with this problem.

VII.

Let me, first, very quickly summarize some of the points I have made. The metaphysical considerations related to Rule 1 seem deep and important, but, as far as I could determine, they provide at best a very nebulous rationale for our division practices, especially with respect to individuation. The middle rules 2 through 4, relating to epistemology, language learning, and explanation, appear on examination to contribute little independently of Rule 1. Finally, the pragmatic considerations last discussed appear to have definite force only in the form of 5.11, and one may doubt whether that principle has either sufficient depth or breadth to provide the sort of rationale we are seeking.

These results are disappointingly negative. At the outset, I introduced a list of division questions featuring some relatively modest possibilities (for example, the possibility of having a word for fast animals or not having a word for puppies). With respect to such examples, one might have been somewhat prepared to believe that there is no serious rationale for our language being the way it is. But, given the strength of our intuitions with respect to the strange languages, one would surely have expected to find clear and decisive arguments showing the absurdity of these languages. Such arguments turn out to be surprisingly elusive. Yet the view that there are essentially no rational constraints on our division practices, that we (or others in the same context) might just as well have spoken even the strange languages, will, I think, continue to seem to most people so farfetched as to be scarcely entertainable. This is a perplexing situation.

One kind of response to this might be in the following vein: "Sometimes arguments that are individually weak can add up to a strong argument. The trouble with the strange languages is that too many connections are broken. In ordinary language there are characteristic connections to similarities and continuities, to projection and explanation, to our interests and concerns. The 'rules' may indeed illuminate what some of these essential connections are. In the strange languages most of these connections are broken or severely warped. Perhaps the question is really not so much whether these are bad languages as whether they are languages at all."

But the puzzle arises from the fact that they do seem to be languages despite all the "broken connections." We could start speaking one of these strange languages tomorrow if we wished, though haltingly and derivatively. Why could there not be people who speak it fluently as their first language? Holistic hand waving about "too many broken connections" does not really seem to put this question to rest.

Someone might answer: "Hand waving is not always bad. One cannot demand more precision than a topic allows, as Aristotle said. These 'strange languages' are obviously insane and idiotic but one may not be able to measure out the ingredients of idiocy and insanity in precise spoonfuls. So one says a little of this and a little of that in the hope of gesturing in the right direction. This may have to suffice."

Perhaps so, but one can try for more. The first step is to recognize the problem.

ELI HIRSCH

Brandeis University

THE JOURNAL OF PHILOSOPHY Digitized by Arya Samaj Foundation Chennai and eGangotri

COMMENTS AND CRITICISM

A NOTE ON CATEGORICAL PROPERTIES AND CONTINGENT IDENTITY

TEPHEN YABLO¹ has attempted recently to revive the notion of contingent identity, identifying this with a relation of coincidence between objects that are "distinct by nature but the same in the circumstances" (296). Yablo argues convincingly for the need of essentialist metaphysics to recognize some relation of this sort, a relation of "intimate identity-like connections between things" (296) if it is to acknowledge properly the intuitive difference between (i) the nonidentity of a bust B and a hunk of wax H of which it is composed, and (ii) the nonidentity of the hunk H and the Treaty of Versailles. (i) and (ii) are clearly not on the same level. Even though B, like the Treaty of Versailles, fails to be strictly identical to H, it is very closely, and quite specially, related to it. What this relation is is certainly worth a general inquiry.²

The suggestion at which Yablo arrives, however, seems to me not adequate to at least certain of the central cases of near identity which motivate his search. Yablo proposes that contingent identity, the desired identity-like connection, is just categorical indiscernibility, that is, possession of all and only the same categorical properties. And categorical properties are properties a "thing has entirely in virtue of how matters actually stand with it" (305). Yablo supplements this characterization of categorical property by offering examples of noncategorical, i.e., hypothetical, properties, ones "grounded not just in how a thing actually is but on how it would or could have been if circumstances had been different" (305). Hypothetical properties include modal, probabilistic, counterfactual, and causal ones.³

But consider Q, a moderately well-known nineteenth-century artist, and Q's fired clay sculpture, W. 4 W, as we know, is not identical with the mass or hunk of clay, M, out of which it is composed, and yet these entities should turn out to be contingently identical if any two things are. This means, according to Yablo's suggestion, that they

¹ "Identity, Essence, and Indiscernibility", this JOURNAL, LXXXIV, 6 (June 1987): 293–314.

² Which is not to agree that resurrecting the label "contingent identity" for it is a wise move, after the powerful attacks of Saul Kripke and others on what is usually meant by that label.

⁸ It is unclear from Yablo's discussion whether he means this to be an exhaustive (or nearly exhaustive) list of kinds of hypothetical property.

⁴ I have switched from wax to clay, to allow for a bit more realism in the subsequent description. Obviously, nothing crucial depends on this.

should be indiscernible with respect to their categorical properties. The problem is that, in reviewing the totality of W's properties, and focusing on those which are not shared with M, I find I am unable to see how they can all be comfortably assigned to the noncategorical bin.

In particular, I have in mind what could be loosely called artistic (or cultural) properties. Firstly, W counts as a work of Q—is in fact one of his proudest accomplishments—whereas M, it appears, is not the work of any artist. Secondly, W belongs to some artistic genre, G—as it happens, that of caricature busts—whereas M belongs to no genre whatsoever. Thirdly, W can be said to exhibit a specific style, reminiscent of Daumier's, whereas M, a hunk tout court, is not in any style. Fourthly, W possesses a proper title, to wit, An Apothecary, yet M, in company with all non-noble nonartworks, lacks a title. And, fifthly, W is, let us grant, innovative in its genre, distinguishing itself from all its predecessors in some technical or expressive respect; M, on the other hand, not being artistic at all, is certainly not artistically innovative.

Now, as far as I can make out, being a work by Q, being a member of a historically constituted genre G, having a certain Daumieresque style, possessing a given title, and being artistically innovative⁸ are properties a "thing has *entirely* in virtue of how things stand with it," properties "grounded entirely in how it is in the circumstances that happen to obtain" (305). Art historical facts of this ilk are the very stuff of actuality, with no apparent admixture of what would or could have been. No causal potential, probabilistic propensity, hidden dis-

Yablo explicitly posits, for the sake of argument, that the temporal durations of wax hunk and sculptural bust in his initial example be precisely the same, i.e., that the hunk and bust come into existence and go out of existence together (294). His subsequent analysis is thus protected against more obvious, purely temporal, counterexamples as the following: W was made in 1888, in P's studio, whereas M was made in 1887, in an argile factory, by some parceler of clay. (I am assuming that being made in 1888 would be a categorical property.) But this protection is purchased at rather high cost, since the assumption required is so counter—realistic. It is scarcely intelligible that a sculpture, a human artistic artifact, and a hunk, an unstructured portion of matter, could have the same beginning of existence in any natural way. So what we thought would end up being explained—how sculptures as normally understood, and the material portions of which they are composed, though strictly nonidentical, are really "at base the same thing"—is not being explained after all, but at most only "basic sameness" between the one and certain stages (or time slices) of the other.

⁶ For more on the metaphysical implications of titling, see my "Titles," Journal of

Aesthetics & Art Criticism, XLIV, 1 (Fall 1985): 29–39.

⁷ I am purposely steering clear of paradigmatic aesthetic properties per se, such as a sculpture's anxiousness, dynamism, or unity, since here clearly a dispositional, and thus categorical-status-disqualifying, analysis notably beckons. (Which is not to admit that such an analysis would be correct.)

⁸ Not to be confused, especially for the issue at hand, with being artistically influential

position, or unrealized possibility appears to lurk in the straightforward conditions of being in the oeuvre of a certain individual, belonging to a particular artistic genre, possessing a given style, having a proper title, or even being stylistically distinct from all other works in its class. Nothing in these properties seems hypothetical, seems to have anything to do with the behavior of their possessors in other possible worlds. Now, clearly, these properties of the sculpture are relational, and thus nonintrinsic, ones. But there is nothing in Yablo's initial (or subsequent) characterization of categorical properties to suggest that categorical properties are restricted to intrinsic ones. ¹⁰

This is not meant to suggest that it is beyond my imagining how one might try to show that the cultural properties of the sculpture I have drawn attention to are in some way covertly hypothetical. But it is to suggest that any maneuvers I am able to imagine would have the consequence that no relational properties (e.g., being tall, having been touched by Jesus) could count as categorical, surely not the desired result.

Nor is what I have said so far meant to deny, as certain other philosophers have done, that sculptures such as the one I described are physical objects, broadly speaking.¹¹ After all, it would be paradoxical, or at any rate implausible, to deny that something composed entirely of material stuff (and not just realized by it, as a program is by hardware), and possessing a definite spatial location to boot, was itself nonmaterial. A sculpture is an unusual kind of physical object, however, one for which a certain configuration, a certain appearance, a certain creative history, and a certain projected purpose are essential. A sculpture is a hunk of material configured in a certain way and governed or structured by a certain intention in the making, namely, the intention that it be regarded-as-a-work-of-art.¹² So con-

⁹ Where, we have often been informed, otherwise unprepossessing objects cavort in all sorts of flambovant fashions.

¹⁰ For example, in discussing the distinction between the Cloth of Turin and the Shroud of Turin, Yablo mentions the (obviously relational) property of having served as the burial shroud of Jesus, and seems to regard this as categorical. (The main difference between the two Turinese entities being that the latter possesses this property necessarily whereas the former possesses it only contingently—a good example of a modal, and thus noncategorical, difference between coincident, i.e., contingently identical, entities. The point of this note, however, is to suggest that it does not appear to work that way always.)

involving carving and molding by hand, as opposed to type sculptures, involving casting or other similar process. The latter are multiple entities, not particulars, and require a different analysis

¹² See my "Defining Art Historically," British Journal of Aesthetics, XIX, 3 (Summer 1979): 232–250; and "Refining Art Historically," Journal of Aesthetics & Art Criticism (Fall 1988).

ceived, it is perforce an object imbued with, and defined by, a cultural identity. As a result, the physical object that is the sculpture—roughly, the hunk-as-so-configured-and-intentioned—has some relatively unusual (viz. cultural) properties, ones of a sort not possessed by its more uncomplicated, physically indiscernible, companion in matter. The sculpture is an intentioned object, as are perhaps all artifacts in a broad sense, yet it is still fundamentally a physical one.¹³

It is presumably not open for Yablo, faced with the five ostensibly categorical artistic properties of the sculpture which I have highlighted, to attempt to appropriate them simply for the clay hunk, even suitably temporally restricted. That is to say, it would be wrong simply to claim that the *hunk* is a work of *P*'s, has a style, is innovative, and so on, even accidentally. The hunk is, admittedly, a different entity, one that composes the sculpture, so the most that could be said is that it composes an object that is a work of *P*'s, has a style, etc.

To be fair to Yablo, he does offer a formal definition of categorical property later in his paper, which is as follows:

. . . a property P is categorical iff necessarily, if a and b are related by refinement, then a has P iff b does . . . (310).

One thing is a refinement of another, for Yablo, just if, in comparing their essences (sets of essential properties) construed in a certain way, it turns out that one is strictly included in the other. So categorical properties are construed as those which will not distinguish between objects, such as certain pairs of near identicals, which stand in a relation of refinement. On the basis of this definition, Yablo is able to show formally that the categorical properties of a thing are precisely the cumulative ones, characterized earlier as those which can contribute to an object's essence without thereby strongly excluding other properties from belonging to it. The virtue of cumulative properties—roughly, purely characterizing ones—is that the essences of different things conceived strictly cumulatively (this is the "certain way" mentioned above) will be directly comparable, allowing, in particular, for essence inclusion relations that are the heart of the idea of one thing being a refinement of another.

¹³ For a related discussion, see my "Zemach on Paintings," *British Journal of Aesthetics*, XXVII, 3 (Summer 1987): 278–283.

¹⁴ This automatically eliminates kind properties, which directly produce incomparability of this sort. (If things are of different basic kinds, then the kind property of the one cannot figure in the complete essence of the other.) It could certainly be clearer, though, what the scope of nontrivial exclusiveness, the mark of a restrictive or noncumulative property, is (299). For example, if a thing's essence contained being (wholly) red this would prevent being green, being yellow, and so on, from being added to this essence. But color properties presumably count as cumulative if any do.

Now, it is true that, according to the conception embodied in this set of definitions, being a work of Q's, being a caricature bust, having a title, and so on, will not be categorical properties—they will not be such as to attach impartially to any two things related by refinement, such as W and M. But this is, at base, because they are not cumulative, seeing as they are all parasitic on a general kind property, that of being a work of art. 15 Yablo's definition of categorical property, it seems, is directly motivated by the desire to have the categoricals come out as precisely the cumulatives, but the connection to the idea of categoricals as properties that things have in a manner independent of what goes on in worlds other than the actual gives the ap-

pearance of having evaporated.

Thus, not only does Yablo not settle the matter, there seems at this point reason to think, in light of my central example, that there is something amiss in his explicit definition of categorical property, or in his preparatory demonstration that categoricality and cumulativeness are, through the refinement relation, necessarily connected,16 or, more broadly, in his attempt to assimilate categorical properties to cumulative ones. For such assimilation seems not to do justice to many intuitively clearcut examples of the categorical, and does not appear to capture the informally expressed idea that categorical properties are to be all those which depend only on how a thing actually is. I can find no error in Yablo's formal demonstration of the equivalence of the categorical and the cumulative as eventually defined, but I remain unconvinced that categoricality and cumulativeness, as each is originally introduced, can turn out to be coextensive. I suspect that it is rather the latter, and not the former, that must ultimately give content to the elusive notion of contingent identity.

Thus, unless this difficulty in the notion or scope of categorical property can be cleared up, even paradigm cases of the elusive "identity-like" relation in question, e.g., that between sculpture Wand hunk M, will not have been accounted for, and Yablo's project of "understanding contingent identity as sameness of nonhypothetical properties" (309) will not have wholly succeeded. There just seem to be "contingent identicals" of a familiar sort which are not equivalent with respect to all categorical properties, intuitively understood, though they perhaps are with respect to all cumulative ones—assuming the latter can be adequately made out.

University of Maryland

JERROLD LEVINSON

¹⁵ See previous note.

¹⁶ See his fn. 19, p. 308.

NOTES AND NEWS

The editors report with sorrow the death of their colleague James Gutman, Professor Emeritus of Philosophy at Columbia University and frequent contributor to this JOURNAL, especially in its earlier years. Professor Gutman taught in the Philosophy Department at Columbia from 1920 until his retirement in 1962, and was a leader in the Ethical Culture Society. He was chairman of the Philosophy Department from 1953 to 1960 and head of the Columbia University seminars from 1969 to 1976. He died on November 6, 1988, at the age of 91. A memorial service is planned in the student lounge, Philosophy Hall, Columbia University, on December 17; contributions may be made to the James Gutman Fund, Columbia University, which was arranged on his ninetieth birthday.

The editors report with sorrow the death of Professor Emeritus Everett J. Nelson, on September 29, 1988. After serving as a faculty member and sometime Chair of the University of Washington Philosophy Department from 1930–1952, Professor Nelson joined the Ohio State Department, and remained there until 1970. During his long academic career, Professor Nelson achieved many honors, including election as President of the Pacific Division of the APA in 1946, and as President of the Western (now Central) Division in 1966.

The editors report with sorrow the death of Max William Wilson, Graduate Professor of Philosophy at Howard University, on April 24, 1988. Professor Wilson was, among other things, a co-founder of the Black Caucus of the American Philosophical Association. Contributions in his memory may be directed to William A. Sadler, Graduate School of Arts & Sciences, Howard Univ., Washington, DC 20059, where a memorial library collection is in the process of being established.

The American Society for Eighteenth Century Studies is pleased to announce the James L. Clifford Prize for an outstanding study of some aspect of eighteenth century culture relevant to any discipline. Articles which have appeared in print in a journal, festschrift, or other serial publication between July 1987 and June 1988 may be nominated for the award by a member of the Society or by its author, by February 1, 1989. Nominations and inquiries may be directed to R. G. Peterson, St. Olaf College, Northfield, MN 55057.

The German Studies Association will hold its thirteenth annual conference in Milwaukee, Wisconsin on October 5–8, 1989. The program committee invites proposals on any topic in German studies, including history, Germanistik, political science, sociology, philosophy, pedagogy, and the arts. Proposals for entire sessions and for interdisciplinary presentations are encouraged. The deadline for proposals is February 15, 1989. Further information may

be obtained by writing Eric Kohler, History Dept., Box 3198, Univ. of Wyoming, Laramie, WY 82071.

The National Endowment for the Humanities would like to announce the 1989/90 Summer Seminars for College Teachers Competitions. In the summer of 1989, sixty-four seminars will be offered on such topics as American cultural and urban history; English Romantic literature, philosophical and anthropological approaches to law, Beethoven's string quartets, Gothic architecture, African history and literature, American constitutionalism, and the influence of Buddhism in China and Japan. Applications are invited from prospective participants for 1989, as well as from prospective Seminar directors for 1990. Proposed topics should focus on enduring issues or current scholarship in the humanities. Scholars wishing to discuss their seminar ideas with staff should submit a draft well in advance of the March 1, 1989 deadline. Requests for further information may be addressed to Summer Seminars for College Teachers, Room 316, Div. of Fellowships & Seminars, National Endowment for the Humanities, 1100 Pennsylvania Ave., Washington, DC 20506.

The Society for the Philosophy of Sex and Love announces a call for papers to be presented at the following two meetings: with the American Philosophical Association, Eastern Division, December 1989; and with the APA Central Division, April 1990. Papers may be on any issue within the philosophy of love and sex. The deadline for the Eastern Division meeting is April 15, and for the Central Division, September 1. Further information may be obtained by writing Alan Soble, Phil. Dept., Univ. of New Orleans, LA 70148.

Harvard University announces the Andrew W. Mellon Faculty Fellowships in the Humanities, offering a one-year appointment, July 1989–1990, with limited teaching duties, departmental affiliation, and opportunity to develop scholarly research. The fellowship is offered to nontenured, experienced junior scholars who have completed, at the time of appointment, two years postdoctoral teaching as college or university faculty in the humanities. Special consideration will be given to candidates who have not recently had access to the resources of a major research university. Inquiries may be directed to Richard Hunt, Harvard University Mellon Faculty Fellowships, 202 Lamont Library, Cambridge, MA 02138.

The Department of Philosophy at Vanderbilt University will host a conference on Descartes to be held February 23–24, 1989. Participants will include Edwin Curley (Illinois/Chicago), Dan Garber (Chicago), Michael Hooker (Maryland/Baltimore County), Jeremy Hyman (UCLA), Peter Markie (Missouri), Gareth Matthews (Massachusetts), Alan Nelson (California/Irvine), Calvin Normore (Toronto), and Margaret Wilson (Princeton). Inquiries may be directed to Jeffrey Tlumak, Phil. Dept., Vanderbilt Univ., Nashville, TN 37235.

INDEX TO VOLUME LXXXV Digitized by Arya Samaj Foundation Chennal and eGangotri

NAMES OF CONTRIBUTORS ARE PRINTED IN SMALL CAPITALS

American Philosophical Association, Eastern Division, 85th Annual Meeting. Symposium Papers, 507 (10), 605 (11); Program 571 (10).

Analytic Truths (Logical and—that are not Necessary). EDWARD N. ZALTA. 57

(2).

Apportioning Causal Responsibility. EL-LIOTT SOBER. 303 (6).

Argument and Perception. APA Symposium. 552 (10).

Argument and Perception: The Role of Literature in Moral Inquiry. MICHAEL R. DEPAUL. 552 (10).

BACH, KENT: Burge's New Thought Experiment: Back to the Drawing Room. 88 (2).

Biological Functions and Perceptual Content. MOHAN MATTHEN. 5 (1).

BOORSE, CHRISTOPHER, and ROY A. SOR-ENSEN: Ducking Harm. 115 (3).

BORDO, SUSAN: Feminist Skepticism and the "Maleness" of Philosophy. 619 (11).

BRANDT, R. B.: Moral Theory and Moral Education. 566 (10).

BROWER, BRUCE W.: Virtue Concepts and Ethical Realism. 675 (12).

BURGE, TYLER: Individualism and Self-

knowledge. 649 (11).

Burge's New Thought Experiment: Back to the Drawing Room. KENT BACH. 88 (2).

CALHOUN, CHESHIRE: Justice, Care, Gender Bias. 451 (9).

CAMPBELL, RICHMOND: Moral Justification and Freedom. 192 (4).

Care(Justice,—,Gender Bias). CHESH-IRE CALHOUN. 451 (9).

CARRIER, DAVID: Review of Jacques Derrida: The Truth in Painting. 219 (4).

Categorical Properties (A Note on—and Contingent Identity). JERROLD LEVINSON. 718 (12).

Causal Responsibility (Apportioning—). ELLIOTT SOBER, 303 (6).

CHILD, T. W.: Review of Colin McGinn: Wittgenstein on Meaning. 271 (5).

Comment on McInerney [abstract].

J. CLAUDE EVANS. 617 (11).

Contingent Identity (A Note on Categorical Properties and—). JERROLD LEVINSON. 718 (12).

Coping with Conflict. FREDERIC SCHICK. 362 (7).

Copyright (Statement of—Policy). 54 (1).

COHEN, G. A., and WILL KYMLICKA: Human Nature and Social Change in the Marxist Conception of History. 171 (4).

CRAIG, WILLIAM LANE: Tachyons, Time Travel, and Divine Omniscience. 135 (3).

Danto, Arthur: The Philosophical Disenfranchisement of Art and The State of the Art. Review by ALEX-ANDER NEHAMAS. 214 (4).

DAVIDSON, DAVID: Reply to Burge [abstract]. 664 (11).

DENNETT, DANIEL C.: Review of Jerry Fodor: Psychosemantics. 384 (7).

DEPAUL, MICHAEL R.: Argument and Perception: The Role of Literature in Moral Inquiry. 552 (10).

DERRIDA, JACQUES: The Politics of Friendship. 632 (11). The Truth in Painting. Review by DAVID CARRIER. 219 (4).

Divine Omniscience. (Tachyons, Time Travel, and—). WILLIAM LANE CRAIG. 135 (3).

DRETSKE, FRED: Review of Alvin Goldman: Epistemology and Cognition. 265 (5).

Ducking Harm. CHRISTOPHER BOORSE and ROY A. SORENSEN. 115 (3).

Ethical Realism (Virtue Concepts and—). BRUCE W. BROWER. 675 (12).

Ethical Universalism and Particularism. ALAN GEWIRTH. 283 (6).

Evans, Gareth. The Varieties of Reference. Review by STEPHEN SCHIFFER. 33 (1).

EVANS, J. CLAUDE: Comments on McInerney [abstract]. 617 (11).

Experiment and Theory: Constitution and Reality. PATRICK A. HEELAN. 515 (10).

Experimentation (The Philosophical Significance of—). APA Symposium. 507 (10).

Explanation of Retribution (An—). ANDREW OLDENQUIST. 464 (9).

Feeling Angry and Elated (On—). STE-PHEN R. LEIGHTON. 253 (5).

Feminist Skepticism. APA Symposium. 619 (11).

Feminist Skepticism and the "Maleness" of Philosophy. SUSAN BORDO. 619 (11).

Fodor, Jerry: *Psychosemantics*. Review by DANIEL C. DENNETT. 384 (7).

Forbes, Graeme: The Metaphysics of Modality. Review by STEPHEN YABLO. 329 (6).

Frankfurt's Concept of Free Will (Sanctification, Hardening of the Heart, and—). ELEONORE STUMP. 395 (8).

FREDE, MICHAEL: The History of Philosophy as a Discipline. 666 (11).

Free Riding and the Prisoner's Dilemma. RAIMO TUOMELA. 421 (8).

Free Will (Sanctification, Hardening of the Heart, and Frankfurt's Concept of—). ELEONORE STUMP. 395 (8).

Freedom (Moral Justification and —). RICHMOND CAMPBELL. 171 (4).

Friedman, Michael: Foundations of Space-time Theories: Relativistic Physics and Philosophy of Science. Review by LAWRENCE SKLAR. 158 (3).

Friendship (The Politics of—). JACQUES DERRIDA. 632 (11).

Further Thoughts on a "Theoretics of Heterogeneity" [abstract]. NAOMI SCHEMAN. 630 (11).

GALISON, PETER: Philosophy in the Laboratory [abstract]. 525 (10).

GASCHÉ, RODOLPHE: Postmodernism and Rationality. 528 (10).

Gender Bias (Justice, Care,—). CHESH-IRE CALHOUN. 451 (9).

Gene (The Return of the—). KIM STER-ELNY and PHILIP KITCHER. 339 (7).

GEWIRTH, ALAN. Ethical Universalism and Particularism. 283 (6).

Goldman, Alvin: Epistemology and Cognition. Review by FRED DRETSKE. 265 (5).

GREEN, EDWARD J.: Review of Frederic Schick: Having Reasons: An Essay on Rationality and Sociality. 28 (1).

GRISWOLD, CHARLES L., JR.: Unifying Plato [abstract]. 550 (10).

HACKING, IAN: On the Stability of the Laboratory Sciences. 507 (10).

Hahn, Lewis Edwin, and Paul Arthur Schilpp, eds.: The Philosophy of W. V. O. Quine. Review by PETER HYLTON. 164 (3).

Harm (Ducking—). CHRISTOPHER BOORSE and ROY A. SORENSEN. 115 (3).

Harrison, Ross: Bentham. Review by David Lyons. 154 (3).

HEELAN, PATRICK A.: Experimentation and Theory: Constitution and Reality. 515 (10).

HIRSCH, ELI: Rules for a Good Language. 694 (12).

History of Philosophy as Actual Philosophy (The—) [abstract]. KENNETH SCHMITZ. 673 (11).

History of Philosophy as a Discipline. APA Symposium, MICHAEL FREDE. 666 (11).

Human Nature and Social Change in the Marxist Conception of History. G. A. COHEN and WILL KYMLICKA. 171 (4).

Husserl (What Is Still Valuable in —'s Analyses of Inner Time-consciousness). APA Symposium, PETER K. MCINERNEY. 605 (11).

HYLTON, PETER: Review of Lewis Edwin Hahn and Paul Arthur Schilpp, eds.: The Philosophy of W. V. O. Quine. 164 (3).

Identity (A Note on Categorical Properties and Contingent—). Jerrold Levinson. 718 (12).

Identity and Survival (Personal—). JOHN ROBINSON. 319 (6).

Indeterminacy (The Refutation of—). JERROLD J. KATZ. 227 (5).

Individualism and Self-knowledge. TYLER BURGE. 649 (11).

IRWIN, T. H.: Review of Martha Nussbaum: The Fragility of Goodness. 376 (7).

Johnsonian Prize in Philosophy. 55 (1). Justice, Care, Gender Bias. CHESHIRE CALHOUN. 451 (9).

Justification (Moral—and Freedom). RICHMOND CAMPBELL. 192 (4).

Justification (Motive and—). STEPHEN SCOTT. 479 (9).

KAHN, CHARLES H.: Plato's *Charmides* and the Proleptic Reading of Socratic Dialogues. 541 (10).

KATZ, JERROLD J.: The Refutation of Indeterminacy. 227 (5). Cogitations. Review by PALLE YOURGRAU. 500 (9).

KITCHER, PHILIP, and KIM STERELNY: Return of the Gene. 339 (7).

KYMLICKA, WILL, and G. A. COHEN: Human Nature and Social Change in the Marxist Conception of History. 171 (4).

Laboratory (Philosophy in the—) [abstract]. PETER GALISON. 525 (10).

Laboratory Sciences (On the Stability of the—). IAN HACKING. 507 (10).

Language (Rules for a Good—). ELI HIRSCH. 694 (12).

Law and Society. APA Symposium. 632 (11).

LEIGHTON, STEPHEN R.: On Feeling Angry and Elated. 253 (5).

LEVINSON, JERROLD: A Note on Categorical Properties and Contingent Identity. 718 (12).

Lewis, David: On the Plurality of Worlds. Review by WILLIAM G. LYCAN. 42 (1).

Literature (Argument and Perception: The Role of—in Moral Inquiry). MI-CHAEL R. DEPAUL. 552 (10).

Logic (Second-order—Still Gone Wild).
MICHAEL D. RESNIK. 75 (2).

Logical and Analytic Truths that are not Necessary. EDWARD N. ZALTA. 57 (2).

LYCAN, WILLIAM G.: Review of David Lewis: On the Plurality of Worlds. 42 (1).

LYONS, DAVID: Review of Ross Harrison: Bentham. 154 (3).

Marxist Conception of History (Human Nature and Social Change in the—).
G. A. COHEN and WILL KYMLICKA. 171
(4).

MATTHEN, MOHAN: Biological Functions and Perceptual Content. 5 (1).

MCCARTHY, THOMAS A.: On the Margins of Politics [abstract]. 645 (11).

McGinn, Colin: Wittgenstein on Meaning. Review by T. W. CHILD. 271 (5).

MCINERNEY, PETER K.: What Is Still Valuable in Husserl's Analyses of Inner Time-consciousness. 605 (11).

Memo to Authors. 53 (1).

Metaphor without Mainsprings: A Rejoinder to Elgin and Scheffler. JOSEF STERN. 427 (8).

Moral Education (Moral Theory and—) [abstract]. R. B. Brandt. 566 (10).

Moral Inquiry (Argument and Perception: The Role of Literature II.).
MICHAEL R. DEPAUL, 5521(10).

Moral Justification and Freedom RICHMOND CAMPBELL 192 (4).

Moral Theory and Moral Education [abstract]. R. B. Brandt. 566 (10).

Motive and Justification. STEPHEN SCOTT. 479 (9). 10 143

Arthur Dany: The Philosophical Disenfranchisement of Art and The State of the Art. 214 (4).

New Books: 98 (2), 390 (7), 440 (8). New Books: Anthologies: 279 (5), 444

New Books in Paperback: 111 (2).

New Books: Translations and New Editions: 225 (4), 447 (8).

Notes and News: 48 (1), 168 (3), 224 (4), 338 (6), 439 (8), 506 (9), 569 (10), 723 (12).

Nussbaum, Martha: The Fragility of Goodness. Review by T. H. IRWIN. 376 (7).

OLDENQUIST, ANDREW: An Explanation of Retribution. 464 (9).

On Feeling Angry and Elated. STEPHEN R. LEIGHTON. 253 (5).

On the Margins of Politics [abstract]. THOMAS A. MCCARTHY. 645 (11).

On the Stability of the Laboratory Sciences. IAN HACKING. 507 (10).

Particularism (Ethical Universalism and—). Alan Gewirth. 283 (6).

Perception (Argument and—: The Role of Literature in Moral Inquiry). MI-CHAEL R. DEPAUL. 552 (10).

Perceptual Content (Biological Function and—). MOHAN MATTHEN. 5 (1).

Personal Identity and Survival. JOHN ROBINSON. 319 (6).

Philosophical Significance of Experimentation, (The—). APA Symposium. 507 (10).

Philosophy (The History of—as Actual Philosophy) [abstract]. KENNETH SCHMITZ. 673 (11).

Philosophy (The History of—as a Discipline). APA Symposium, MICHAEL FREDE. 666 (11).

Philosophy in the Laboratory [abstract]. PETER GALISON. 525 (10).

Plato and the Propleptic Reading of So-

cratic Dialogues. APA Symposium. 541 (10).

Plato's Charmides and the Proleptic Reading of Socratic Dialogues. CHARLES H. KAHN. 541 (10).

Plato (Unifying—) [abstract]. CHARLES L. GRISWOLD, JR. 550 (10).

Politics of Friendship (The—). JACQUES DERRIDA. 632 (11).

Politics (On the Margins of—) [abstract]. THOMAS A. MCCARTHY. 645 (11).

Postmodernism and Rationality. APA Symposium, RODOLPHE GASCHÉ. 528 (10).

Proudfoot, Wayne: Religious Experience. Review by NINIAN SMART. 151 (3).

Rationality (Postmodernism and—). RODOLPHE GASCHÉ. 528 (10).

Realism (Virtue Concepts and Ethical

—). BRUCE W. BROWER. 675 (12).

Reason (The Question of—) [abstract]. CHARLES SCOTT. 539 (10).

Refutation of Indeterminacy (The—). JERROLD J. KATZ. 227 (5).

Reply to Burge [abstract]. DONALD DA-VIDSON. 664 (11).

RESNIK, MICHAEL D.: Second-order Logic Still Gone Wild. 75 (2).

Responsibility (Apportioning Causal—). ELLIOTT SOBER. 303 (6).

Retribution (An Explanation of—). ANDREW OLDENQUIST. 464 (9).

Return of the Gene, (The—). KIM STER-ELNY and PHILIP KITCHER. 339 (7).

ROBINSON, JOHN. Personal Identity and Survival. 319 (6).

Rules for a Good Language. ELI HIRSCH. 694 (12).

Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will. ELEONORE STUMP. 395 (8).

SCHEMAN, NAOMI: Further Thoughts on a "Theoretics of Heterogeneity" [abstract]. 630 (11).

Schick, Frederic: Coping with Conflict. 362 (7). Having Reasons: An Essay on Rationality and Sociality. Review by EDWARD J. GREEN. 28 (1).

SCHIFFER, STEPHEN: Review of Gareth Evans: The Varieties of Reference. 33 (1).

Schilpp, Paul Arthur, and Lewis Edwin Hahn, eds.: The Philosophy of

W. V. O. Quine. Review by PETER HYLTON. 164 (3).

SCHMITZ, KENNETH: The History of Philosophy as Actual Philosophy. 673 (11).

Sciences (On the Stability of the Laboratory—). IAN HACKING. 507 (10).

SCOTT, CHARLES: The Question of Reason [abstract]. 539 (10).

SCOTT, STEPHEN: Motive and Justification. 479 (9).

Second-order Logic Still Gone Wild. MICHAEL D. RESNIK. 75 (2).

Self-knowledge (Individualism and—). APA Symposium, TYLER BURGE. 649 (11).

Skepticism (Feminist—). APA Symposium. 619 (11).

Skepticism (Feminist—and the "Maleness" of Philosophy). SUSAN BORDO. 619 (11).

SKLAR, LAWRENCE: Review of Michael Friedman: Foundations of Space-time Theories: Relativistic Physics and Philosophy of Science. 158 (3).

SMART, NINIAN: Review of Wayne Proudfoot: Religious Experience. 151 (3).

SOBER, ELLIOTT: Apportioning Causal Responsibility. 303 (6).

Social Change (Human Nature and—in the Marxist Conception of History).
G. A. COHEN and WILL KYMLICKA. 171
(4).

Socratic Dialogues (Plato's Charmides and the Proleptic Reading of—). CHARLES H. KAHN. 541 (10).

SORENSEN, ROY, and CHRISTOPHER BOORSE: Ducking Harm. 115 (3).

STERELNY, KIM, and PHILIP KITCHER: The Return of the Gene. 339 (7).

STERN, JOSEF: Metaphor without Mainsprings: A Rejoinder to Elgin and Scheffler. 427 (8).

STUMP, ELEONORE: Sanctification, Hardening of the Heart, and Frankfurt's Concept of Free Will. 395 (8).

Survival (Personal Identity and—). JOHN ROBINSON. 319 (6).

Tachyons, Time Travel, and Divine Omniscience. WILLIAM LANE CRAIG. 135 (3).

Theory (Experiment and—: Constitution and Reality). PATRICK A. HEELAN. 515 (10).

- Thought Experiment (Burge's New—: Back to the Drawing Room). KENT BACH. 88 (2).
- Time-consciousness (What Is Still Valuable in Husserl's Analyses of Inner—). APA Symposium, PETER K. MCINERNEY. 605 (11).
- Time Travel (Tachyons,—, and Divine Omniscience). WILLIAM LANE CRAIG. 135 (3).
- TUOMELA, RAIMO: Free Riding and the Prisoner's Dilemma. 421 (8).
- Unifying Plato [abstract]. CHARLES L. GRISWOLD, JR. 550 (10).
- Universalism (Ethical—and Particularism). ALAN GEWIRTH. 283 (6).

- Virtue Concepts and Ethical Realism. BRUCE W. BROWER. 675 (12).
- What Is Still Valuable in Husserl's Analyses of Inner Time-consciousness. APA Symposium, PETER K. MCINERNEY. 605 (11).
- YABLO, STEPHEN: Review of Graeme Forbes: The Metaphysics of Modality. 329 (6).
- YOURGRAU, PALLE: Review of Jerrold J. Katz: Cogitations. 500 (9).
- ZALTA, EDWARD N.: Logical and Analytic Truths that are not Necessary. 57 (2).

Ethics of Coercion and Authority A Philosophical Study of Social Life

TIMO AIRAKSINEN

A novel and carefully argued theory of the nature of power and the sources of social conflict in liberal society. Airaksinen argues that power is a philosophical challenge which demands a more rigorous description than standard sociological talk has allowed, and he faces that challenge by considering the twin notions of *coercion*—control based on threats—and *authority*—control based on trust.

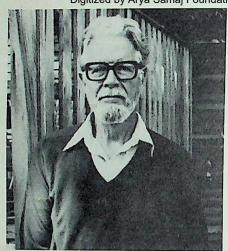
Illustrated with examples from life and literature, this is the first full-scale philosophical treatise to draw a clear analytical distinction between these forms of social control. Airaksinen considers the problems of deterrence, exploitation, the justification of punishment, and, critically, the rationality of accepting authority. The works of John Rawls, David Gauthier, Joseph Raz, Anthony Giddens, and Robert Nozick inform the discussion, as their views on power, rationality, and authority are contrasted with those of Marx and Hegel.

\$25.95

University of Pittsburgh Press

c/o CUP Services, Box 6525, Ithaca, NY 14850 • 800-666-2211

Digitized by Arya Samaj Foundation Chennai and eGangotri



The Company We Keep

An Ethics of Fiction WAYNE C. BOOTH

"Wayne Booth reminds us that a work of narrative literature speaks with a human voice, presents human characters and actions, and calls for a human response that inescapably involves ethical considerations. . . . A splendid exemplar of ethical criticism."

—M. H. Abrams, author of The Mirror and the Lamp: Romantic Theory and the Critical Tradition \$29.95

The Limitations of Deductivism ADOLF GRÜNBAUM and WESLEY C. SALMON, Eds.

These essays,, by distinguished philosophers of science explore the extent to which the scope of deduction is limited, and the ways in which it must be supplemented to produce an adequate account of the methods of science. \$32.50

Empirical Knowledge

"In both conception and execution this is a fine book. . . . Fresh and invigorating." $\hfill \hfill \hfill$

—Frederick Schmitt, University of Illinois at Urbana-Champaign "Novel, powerful and yet fairly simple. . . .persuasive and as well worked-out as any I know. . . .just excellent."
—William Gregory Lycan, University of North Carolina at Chapel Hill \$45.00

Rights in Moral Lives

A Historical-Philosophical Essay A. I. MELDEN

"I believe it to be an exceptionally important book—not a trendy one, but rather one whose importance will be sustained for quite some time." —Allen E. Buchanan, University of Arizona \$25.00

Weber, Irrationality, and Social Order

ALAN SICA

"Sica has given us a book of profound scholarship and theoretical imagination. He raises the reading of Weber to a new level of seriousness and sophistication."

-Robert Bellah, University of California, Berkeley \$32.50

ζS

CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

Digitized by Arya Samaj Foundation Chennai and eGangotri



Digitized by Arya Samaj Foundation Chennai and eGangotri CC-0. In Public Domain. Gurukul Kangri Collection, Haridwar

